BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

MICROBIOLOGY AND MOLECULAR BIOLOGY REVIEWS, Mar. 2000, p. 153–179
1092-2172/00/$04.00+0

Vol. 64, No. 1

Microbial Relatives of the Seed Storage Proteins of Higher Plants: Conservation of Structure and Diversification of Function during Evolution of the Cupin Superfamily

JIM M. DUNWELL,[1]* SAWSAN KHURI,[1] AND PAUL J. GANE[2]

School of Plant Sciences, The University of Reading, Reading,[1] and Drug Design Group, Department of Biochemistry, University of Cambridge, Cambridge,[2] United Kingdom

* Corresponding author. Mailing address: School of Plant Sciences, The University of Reading, Whiteknights P.O. Box 221, Reading RG6 6AS, United Kingdom. Phone: 44-118-931-6313. Fax: 44-118-931-6577. E-mail: J.M.Dunwell@reading.ac.uk.

INTRODUCTION

The recent publication of the sequences of several complete nomes of archaea and bacteria has stimulated a range of new alyses of gene and protein evolution. These studies have cluded many which have considered the distribution of specific families of paralogs (families of related proteins from the ame species) and orthologs (families of related proteins from ifferent species). The power of these analyses (mostly dependent on algorithms designed to detect similarities in gene or rotein sequences) lies in their ability to identify similarity in he many million sequences now held in the major databases. However, despite the undoubted efficiency of these comparative studies, there remain several constraints, which limit the value of any new information that can be generated. First, each algorithm depends upon a certain level of similarity (usually above 30% identity) to detect a statistically valid relationship between two or more sequences. It is much more difficult, though not impossible, to confirm similarity where the degree of identity between sequences is 20% or lower. Second, simple analysis of primary sequence provides no information about the secondary or tertiary structure of the protein(s) under investigation, and it is the structure of a protein that determines its function. There is therefore a growing interest expanding from genome and transcriptome analysis (299) into structural genomics (14) and studies of the proteome and metabolome present in any specific cell or tissue (89, 159, 271).

This present review is designed to show how a detailed analysis of protein sequence has been combined with information on tertiary structure and biochemical function to uncover a new superfamily of functionally diverse proteins, the cupins, and to trace their evolution from bacteria and archaea to eukaryotes including animals and higher plants. Specifically, this path leads from small enzymes found in primitive thermophilic microbes to plant enzymes of great medical value and thence to the multimeric seed storage proteins that comprise the major part of the human diet.

DEFINITION OF THE CUPIN SUPERFAMILY

The term cupin (from the Latin term "*cupa*," for a small barrel or cask) has been given (64) to a β-barrel structural domain identified in a superfamily of prokaryotic and eukaryotic proteins that include several enzymes, as well as factors that bind sugars and other compounds (69). This superfamily also includes many of the storage proteins from higher plants (20), and it was the knowledge of the three-dimensional structures of these proteins (155, 177) that allowed the molecular modelling of the wheat protein germin (90), an unusual protease-resistant protein with oxalate oxidase (OXO) (EC 1.2.3.4) activity (173). The main characteristic of the cupin domain is a two-motif sequence (69) in which motif 1 corresponds to the C and D strands and motif 2 corresponds to the G and H strands of the unit structure of the bean storage protein phaseolin (177). Between these two motifs (usually His containing) is a region, containing strands E and F, that varies in length from 15 residues in many of the bacterial enzymes to more than 50 residues in some of the storage proteins (see Fig. 1); the exact number of residues is one diagnostic feature of each subclass of protein. The other main diagnostic feature is the overall organization of the protein, which can comprise either a single domain, as in the germin and germin-like proteins (46, 200), or a duplicated, two-domain structure. This latter structure was identified first in the storage proteins and was considered to be part of a presumed evolutionary progression from a single-domain, eukaryotic precursor (20). It now seems possible that the critical duplication event actually occurred in a prokaryote, with subsequent evolution leading to the two-domain proteins in higher plants. For example, two such duplicated proteins, one from the cyanobacterium *Synechocystis* and one from the gram-positive bacterium *Bacillus subtilis*, were identified in 1998 by Dunwell and Gane (69), who also described a similar two-domain composition in an oxalate decarboxylase (OXDC) (EC 4.1.1.2) from the wood-rotting fungus *Collybia velutipes* (now termed *Flammulina velutipes*). On the basis of these discoveries, Dunwell and Gane proposed the hypothesis that all the higher-plant storage proteins, the major component of the human diet, evolved from such duplicated, microbial sequences. It now seems much more likely (260) that the particular duplication event leading to the storage proteins in higher plants occurred independently of that producing the fungal OXDC enzymes.

In this review, the individual members of the cupin superfamily are described in terms of their primary amino acid sequence, in addition to their structure and function (where these are known). Particular attention is given to a detailed analysis of the cupin gene family in *B. subtilis*, the prokaryote with the most complete range of relevant sequences described to date. Finally, an assessment will be made of the biological significance of various cupins, the present practical value of some cupin microbial enzymes used in medicine, agriculture, and industry, and some possible future research directions.

ANALYTICAL METHODS USED TO IDENTIFY CUPIN SEQUENCES

The original starting point for this analysis was the identification of the so-called germin box (171), a nonapeptide sequence (HI/THPRATEI) found in both the two wheat germins (GF-2.8 and GF-3.8) and the spherulins, a group of proteins produced during encystment of the slime mold *Physarum polycephalum* (29). Previous analysis at PROSITE had designated at PDOC00597 a germin family signature that included

the germin box; in addition, there is a three-element PRINTS fingerprint GERMIN based on the alignment of 12 proteins. It should be stressed that these prior analyses are now outdated since they used only a small proportion of the currently available sequences. In particular, many of them use data only from the SwissProt database, a source that contains less than one-third of the available data.

The starting point for the secondary stage of this study was the conserved two-motif structure of cupins [conserved motif 1, $PG(X)_5HXH(X)_4E(X)_7G$; conserved motif 2, $G(X)_5PXG(X)_2H(X)_3N$] with a variable intermotif spacing of 15 to ca. 50 amino acids (aa) (69). This two-motif signature is located within several conserved sequences, including ProDom (49) (release 34.1) domains 2426 (this includes germin and germin-like proteins from higher plants), 1428 (derived from bacterial phosphomannose isomerases, GDP mannose-1-phosphate pyrophosphorylases, and polyketide synthases), 45821 (bacterial regulatory proteins), and 6286 (bacterial AraC-type transcription factors). Presumably, the conserved cupin motifs within these domains had not been identified previously because of the varied intermotif spacing.

The present systematic analysis was initiated using the individual cupin motifs referred to above, together with sequences spanning the two motifs. A series of iterative database searches was conducted using the gapped Blast (7) and BLOCKS (114, 225) programmes. The principal database used for this analysis was the nonredundant GenBank site maintained at the National Institute for Biotechnology Information, National Institutes of Health, Bethesda, Md., but analyses were also conducted on the genome of *B. subtilis* at SubtiList (http://www.pasteur.fr/Bio/SubtiList.html) and the genome of *Synechocystis* (145, 146), at CyanoBase (http://www.kazusa.or.jp/cyano/scarch/html) (211). Other microbial genomes, either complete or unfinished, were accessible via GenBank and other sites, including those at NIH (http://www.ncbi.nlm.nih.gov/BLAST/unfinishedgenome.html), the Institute of Genome Research (Rockville, Md.) (http://www.tigr.org/tigr-home/tdb/mdb.html), and the Sanger Centre (Cambridge, United Kingdom) (http://sanger.ac.uk/).

To identify previously unknown cryptic coding regions and their protein products (see below), particular attention was paid to TBlastN searches. In many cases, these searches revealed significant matches in more than one reading frame (ORF) from a single gene (or expressed sequence tag [EST]) sequence. This suggested the likelihood of insertions or deletions in the DNA sequence as a consequence of cloning or sequencing errors. Manual editing was therefore conducted on such sequences to generate amended polypeptide sequences, which were then tested in further searches. Alignments of proteins and DNA sequences were conducted using a variety of programmes including Clustal, MAP, Pima, and GeneQuiz (http://columba.ebi.ac.uk:8765/ext-genequiz/).

One specific aim of the present study was to conduct a detailed analysis of information from whole-genome sequencing projects (42, 48, 53, 79, 80, 147, 153, 268), particularly that of the gram-positive bacterium *B. subtilis* (166), in order to assess more completely the range of cupin sequences in this bacterium and to confirm its identity as the most likely progenitor of the spectrum of cupins found in higher plants. The results of this analysis are reported below.

MEMBERS OF THE CUPIN SUPERFAMILY

The following sections provide details of each subclass of the cupin superfamily in turn, first categorized according to the primary protein sequence (i.e., simple structure with a single cupin domain, complex structure with a single domain, or duplicated structure with two cupin domains) and then categorized according to the number of residues between the two conserved motifs present within each domain. Figure 1 provides an alignment of a selection of putative cupin sequences arranged to show the two conserved motifs together with the increase in intermotif spacing from the basic value of 15 in many microbial enzymes up to 54, as found in a representative storage protein. It is acknowledged that absolute confirmation that all these sequences belong to the cupin family must await resolution of their tertiary structure, but in the meantime it is reasonable to propose this as a working hypothesis—an approach supported by an independent study (12) using PSI-BLAST (7).

SINGLE-DOMAIN CUPINS

The great majority of cupin proteins contain only a single conserved domain at the core of the protein. Within this large grouping, the various subclasses considered below can be categorized not only on the basis of the variable intermotif spacing within this domain but also on the basis of the specific conserved residues within each motif and, to a lesser extent, within the intermotif region. In the great majority of examples, the first motif comprises 20 or 21 residues and the second motif has 16 residues (Fig. 1). The minimum intermotif spacing found in cupins is 15 residues; this includes strands E and F together with the interstrand loop. Presumably, there are steric constraints in the tertiary structure that do not permit a shorter loop. Analysis from the various genome-sequencing projects (J. M. Dunwell, unpublished data) has now revealed a total of more than 200 microbial sequences with this 15-residue spacing.

Phosphomannose Isomerases

Phosphomannose isomerases (PMI) (EC 5.3.1.8) are enzymes that catalyze the interconversion of mannose-6-phosphate and fructose-6-phosphate. The subclass most relevant to this review is that of the type II enzymes (139, 227), known to be involved in a variety of microbial pathways including capsular polysaccharide biosynthesis and D-mannose metabolism. Such enzymes, which contain the two-motif cupin signature separated by 15 aa, exist either as a single-function protein of about 120 to 150 aa or as the C-terminal domain of a bifunctional enzyme (ca. 480 aa) with both PMI and GDP-mannose pyrophosphorylase (GMP) (EC 2.7.7.22) activity. An example of the latter type of protein, and one of particular practical importance, is the 56-kDa bifunctional enzyme encoded by *algA* (179, 196, 259), which catalyzes the first and third steps in the biosynthesis of alginate, PMI catalyzing the first step (152). This compound is composed of 1,4-linked α-L-guluronic acid and β-D-mannuronic acid and is of great economic importance, although for commercial production it is usually extracted from marine seaweeds rather than from bacteria (237). Alginate also has medical significance because of its production by *Pseudomonas aeruginosa* during the conversion of this bacterium to a mucoid form (256). This conversion is induced by several conditions: starvation, the presence of metabolic inhibitors, or, most importantly, growth of the bacteria in the lungs of cystic fibrosis patients. Indeed, mortality in such patients is usually associated with the inability of antibiotics to penetrate the bacterial biofilm and to the fact that the alginate protects the bacteria from the host immune responses (136). Similarly, alginate is a major component of metabolically dormant cysts in the aerobic nonsymbiotic soil bacterium

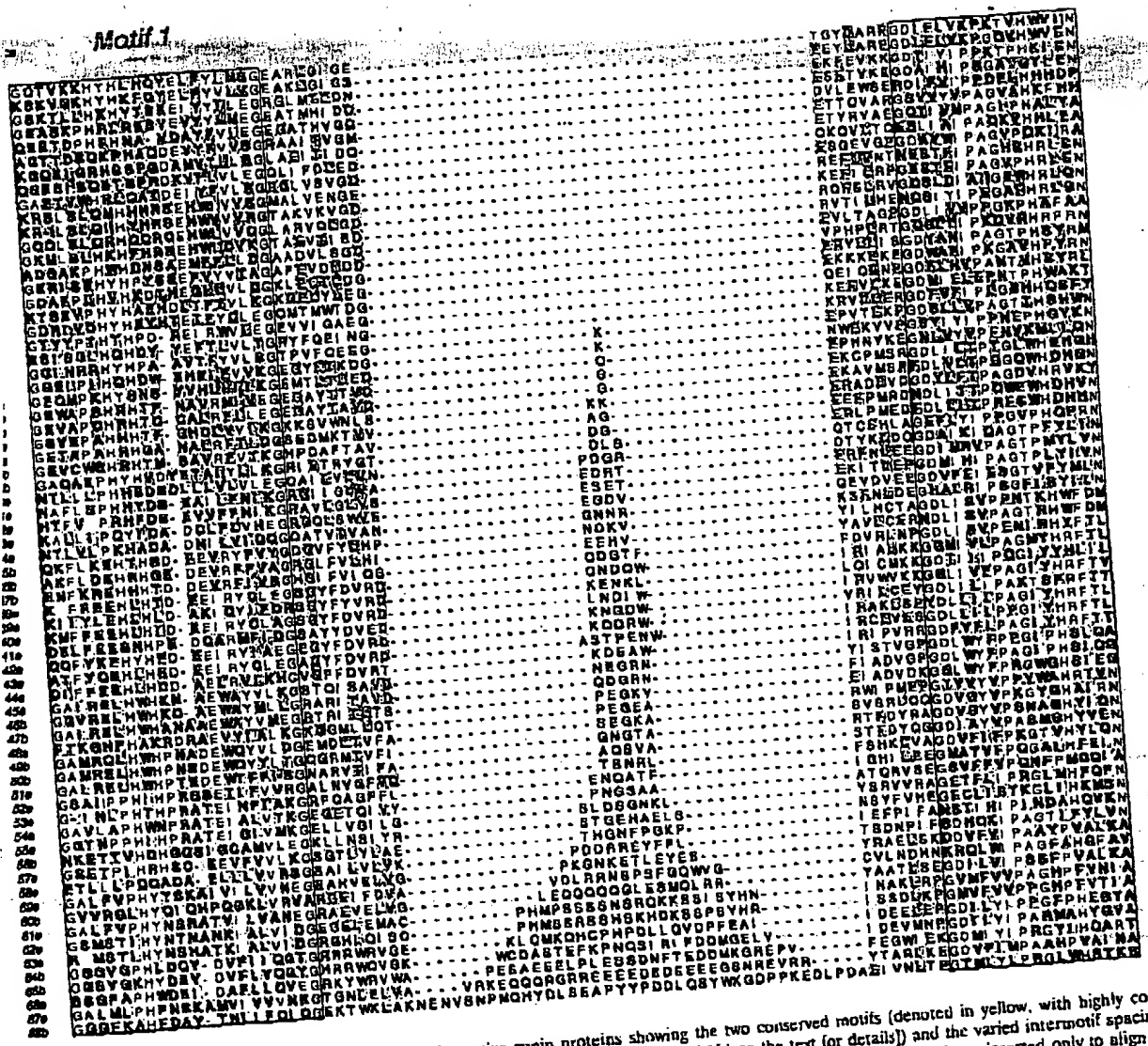


FIG. 1. Multiple alignment of a representative sample of putative cupin proteins showing the two conserved motifs (denoted in yellow, with highly conserved ...idues in red) (motif 1 corresponds to strands C and D; motif 2 corresponds to strands G and H [see the text for details]) and the varied intermotif spacing of 15 ...54 aa. Conserved residues in strands E and F are shaded in grey. In all cases the sequences are continuous and the gaps have been inserted only to align the two ...tifs. The source organism and the GenBank gi identifier (in parentheses) of the sequences are as follows: 1, *Desulfurococcus* sp. (1545809); 2, *Pyrococcus horikoshii* ...256943); 3, *Methanococcus jannaschii* (2128971); 4, *Methanobacterium thermoautotrophicum* (2621410); 5, *Haloferax* sp. strain D1227 (3293533b); 6, *Streptomyces* ...*eolicolor* (5457273); 7, *Streptococcus pyogenes* (contig 7); 8, *S. pyogenes* (contig 112); 9, *Mycobacterium tuberculosis* (2104394); 10, *Pseudomonas aeruginosa* (3510759); ...1, *P. horikoshii* (3131181); 12, *Synechococcus* sp. (79640); 13, *Rhizobium* sp. (2499713); 14, *S. coelicolor* (4467248); 15, *S. halstedii* (730725); 16, *Bacillus subtilis* ...2636545a); 17, *Aquifex aeolicus* (2984227); 18, *B. subtilis* (2636545b); 19, *A. aeolicus* (Table 1); 20, *Escherichia coli* (116101); 21, *Enterobacter aerogenes* (1572541); 22, ...*horikoshii* (3256432); 23, *M. jannaschii* (2833572); 24, *Pseudomonas* sp. strain U2 (4220433a); 25, *Sphingomonas* sp. strain RW5 (3550667a); 26, *Arabidopsis thaliana* ...1169199); 27, *Haloferax* sp. strain D1227 (3293534a); 28, *Nocardioides* sp. (2588983a); 29, *Erwinia chrysanthemi* (1772621); 30, *Canavalia ensiformis* (17977a); 31, *Pisum* ...*ativum* (2765097a); 32, *Glycine max* (548900a); 33, *Matteucia struthiopteris* (1019792a); 34, *Arachis hypogaea* (1168390a); 35, *A. aeolicus* (2984230); 36, *P. aeruginosa* ...contig 54); 37, *B. subtilis* (2633733); 38, *Oryza sativa* (2952338); 39, *H. sapiens* (3201599); 40, *A. thaliana* (2739365); 41, *Osteragia ostenagi* (2996183); 42, *Candida* ...*albicans* (Con4-2749); 43, *Saccharomyces cerevisiae* (2497115); 44, *Caenorhabditis elegans* (3877049); 45, *Collybia velutipes* (1604990a); 46, *B. subtilis* (2634260a); 47. ...*Synechocystis* strain PCC6803 (1652630a); 48, *Pyrococcus horikoshii* (3258400); 49, *Synechocystis* strain PCC6803 (1652630b); 50, *B. subtilis* (2634260b); 51, *C. velutipes* ...(1604990b); 52, *Pinus caribaea* (274548); 53, *Phytanum polycephalum* (134860); 54, *M. struthiopteris* (1019792b); 55, *Triticum aestivum* (121129); 56, *B. subtilis* (2635598); ...57, *A. thaliana* (461453); 58, *Phaseolus vulgaris* (230247a); 59, *P. vulgaris* (230247b); 60, *Synechocystis* PCC6803 (1653678); 61, *Canavalia gladiata* (18007b); 62, *P. sativum* ...(2765097b); 63, *G. max* (548900b); 64, *E. coli* (1787373); 65, *Haemophilus influenzae* (1175655); 66, *C. elegans* (2047349); 67, *A. hypogaea* (1168390b); 68, *B. subtilis* ...(2636534). The suffix a, b, or c following the sequence number (1 through 68) in the figure refers to either the first or second domain, respectively, in a bicupin sequence. The suffix a or b after the gi identifier above refers to the organism as being an archaeon, eubacterium, or eukaryote, respectively.

Azotobacter vinelandii, where it may account for up to 70% of the intine (inner layer of wall) and 40% of the exine (outer layer of wall) carbohydrates. This coating is believed to protect the cell from desiccation and other stresses, and indeed its production in the lungs of cystic fibrosis patients may be linked to the need for the bacterial cells to protect themselves from the dehydrating environment.

The equivalent bifunctional enzyme in *Escherichia coli* is ManC (gi|3435180), part of the biosynthetic pathway for GDP-L-fucose and GDP-perosamine, components of the O-

antigen gene cluster (140, 280, 283, 301). Other related bacterial enzymes include those encoded by *noeJ* from *Rhizobium* (81) and *aceF*, which is part of the acetan biosynthetic pathway in *Acetobacter xylinus* (102). There are also related genes in the archaeal species *Pyrococcus horikoshii* (gi|3257338), *Methanobacterium thermoautotrophicum* (gi|2622642), and *Archaeoglobus fulgidus* (gi|2649495).

Because of its importance in the synthesis of bacterial and fungal cell walls, PMI inhibition is a target for drug discovery (32). Although there is limited information on the structure of the active site, in the context of the conserved histidines in the two cupin motifs it is pertinent to note recent evidence (220) for the existence of a His residue in this site in a PMI from *Xanthomonas campestris*; this particular PMI is considered to be a metalloenzyme and is activated by zinc.

Polyketide Synthases (Putative Cyclases)

The polyketide pathway (115, 125, 194, 235) accounts for the biosynthesis of many of the thousands of known secondary metabolites, including antibiotics and pigments. Among these products is curamycin (26), an antibiotic produced by *Streptomyces curacoi* and based on a polyketide skeleton consisting of a modified orsellinic acid—an unreduced version of 6-methylsalicylic acid and the simplest of all aromatic polyketides. It was found (25) that the gene cluster responsible for the synthesis of this antibiotic was very similar to the *S. coelicolor whiE* gene cluster responsible for the synthesis of a grey spore pigment produced shortly before sporulation in the aerial mycelium (51), and subsequent studies (36) demonstrated the widespread occurrence of gene clusters very similar to *whiE* among other *Streptomyces* spp. Of specific interest to this review is the sequence of the homologous group of genes represented by *curC* (*S. curacoi*), *whiE* ORFII (*S. coelicolor*) (8), *sch* ORFB (*S. halstedii*), and *tcmJ* (*S. glaucescens*) (33). The exact biochemical function of these gene products remains unknown, although it is suggested to be a cyclase (148, 318). Sequence analysis reveals the two conserved cupin motifs, separated by a distance of 15 residues, within a total protein size of approximately 150 aa. It has been suggested recently (12) that use of the CurC sequence is the most efficient means of identifying other members of the cupin family in a PSI-BLAST search (7).

Recent analysis (69) has extended the number of members in this particular cupin subfamily to include several other close relatives, such as the sequence gi|2635101 (YrkC) from *B. subtilis* and the 140-aa Pep1 sequence (gi|1572541) encoded by gene *tnpA* of the cryptic transposon Tn*4321* within the broad-host-range IncPβ plasmid R751 of *Enterobacter aerogenes* (267, 291). The notes accompanying the Pep1 database submission recognized a "possible polyketide cyclase on basis of weak similarity to TcmJ of *S. glaucescens*" (E value 8.5). However, it is most similar (E value 8e-08) to a 97-aa sequence encoded by nucleotides 180 to 467 of a contig (gnl|Stanford_382|smelil_423025B02.xl) from *Sinorhizobium meliloti*.

It was assumed previously that the smallest of all cupins is the 77-aa "membrane-spanning protein" gi|1017816 from *Streptomyces coelicolor* (181, 182). However, the start codon for this sequence has been reassigned, and it is now considered to encode a 115-aa protein (gi|5457273) that is most similar to a 79-aa polypeptide encoded by nucleotides 243111 to 243347 from contig 7 of *Streptococcus pyogenes*.

Dioxygenases

Several types of dioxygenase enzymes are probable members of the cupin superfamily. They can be divided into two categories, those with a single domain and those with two domains (bicupins); within each subcategory the individual members can be recognized on the basis of a characteristic inter-motif spacing.

3-Hydroxyanthranilate 3,4-dioxygenase (3-HAO) (EC 1.13.11.6), with an intermotif spacing of 19 or 23 aa, is a eukaryotic enzyme that cleaves the aromatic ring of 3-hydroxyanthranilic acid to produce 2-amino-3-carboxymuconic semialdehyde, an intermediate in the synthesis of the excitotoxin quinolinic acid (21); this compound kills neurons by activation of *N*-methyl-D-aspartate receptors, and inhibition of 3-HAO is therefore a pharmaceutical target (35). The enzyme is well characterized in mammals (210) and is part of the kynurenine pathway for the catabolism of tryptophan. Recently, the yeast gene *YJR025c* has been shown (164) to encode a 3-HOA (gi|1353060) homologous to the human equivalent (190) and has been renamed *BNA1* (biosynthesis of nicotinic acid). A very similar polypeptide (E value 5e-64) is encoded by part of a contig (gnl|Stanford_5476|C.albicans_Con4-2428) from *Candida albicans*. Alignment of these 3-HOA sequences shows a notable difference between the *Saccharomyces* sequence and the other sequences, in that the former protein has an intermotif spacing of 23 residues compared with 19 for the other sequences. This insertion of 4 aa occurs in the loop between the E and F strands of the barrel.

In common with most other dioxygenase enzymes, 3-HAO requires nonheme iron as a cofactor. However, in contrast to the multimeric composition of the related, two-domain dioxygenases described below, this enzyme seems to be monomeric.

Cysteine dioxygenase (CDO) (EC 1.13.11.20), with an intermotif spacing of 28 residues, is a key enzyme of cysteine metabolism and catalyzes the production of cysteine sulfinate. The rat (296), human (232), and *Caenorhabditis elegans* genes have been well characterized, with the closest bacterial relatives of these eukaryotic sequences being those from *B. subtilis* (gi|2635598), *Streptomyces coelicolor* (gi|2687337), and *Mycobacterium tuberculosis* (gi|2896702). This enzyme is known to be monomeric, with one atom of iron per molecule (312); its activity is strongly reduced by chelators of Cu^{+} and Fe^{2+} (247).

Spherulins

The life cycle of the simple slime mold *Physarum polycephalum* involves a transition between two vegetative states, the amoeba and the plasmodium. Amoebae are the uninucleate haploid cells, which under some conditions will fuse and differentiate into a giant multinucleate diploid plasmodium. When these latter cells are grown in liquid medium, they fragment into microplasmodia, which are capable of withstanding adverse conditions by encystment. This transition into hard-walled oligonucleated spherules is termed spherulation, and it is induced by starvation (or high concentrations of some carbohydrates), cooling, dehydration, acidic pH, and/or sublethal concentrations of heavy metals (47, 144).

As part of a molecular study of this phase transition, it was shown first that the major changes in protein synthesis take place 24 h after the beginning of starvation-induced spherulation (31) and that the four most abundant spherulation-specific RNAs accounted for more than 10% of all mRNAs present after this period (30); these mRNAs were not present in encysting amoebae or in sporulating plasmodia. Differential hybridization of a cDNA library was used subsequently to isolate full-length clones (29), of which two were found to be 76% similar and encoded proteins named spherulins 1a and 1b (81% identical). These proteins possess a potential signal peptide and an N-glycosylation site and were therefore presumed to be cell-wall glycoproteins. It was discovered subsequently

hat there is 44% similarity at the amino acid level be-
spherulin 1b and the wheat germin GF-2.8; this value
ses to 60% for the central core sequence, the region that
ns the conserved PH(I/T)HPRATEI decapeptide desig-
the germin box. They can thus be considered cupins, with
ermotif spacing of 21 aa.

interesting addition to the discussion on the evolutionary
1 of the spherulin genes is provided by an analysis of
n position (20) in a series of related cupins. The discovery
the C-terminal domain of several seed storage proteins
those of *Welwitschia mirabilis* and *Gingko biloba*) shared
ntron position with the spherulins (although shifted by 2 bp
. polycephalum) provided strong support for the concept
t these proteins have a common ancestor.

To date, no biochemical function has been assigned to these
herulins, although they do not seem to have any OXO ac-
ity (173). However, it is relevant to consider their possible
nction(s) in the specific context of what is known about the
nditions pertaining during spherulation and also in the gen-
ral context of the link between cupins and stress responses in
rokaryotes and eukaryotes. In particular, it is interesting to
ote the link between oxidative stress and spherulation. The
nitial circumstantial evidence for such a link came from the
observation (2) that the berbicide paraquat, a compound that
generates free radicals, accelerated spherulation and also in-
creased the specific activity of the manganese isoform of su-
peroxide dismutase (3). It was also found that during the
spherulation process in salts-only starvation medium, superox-
ide dismutase activity increased 46-fold, along with an increase
in the concentrations of H_2O_2 and organic peroxide (2); none
of these changes occurred in nondifferentiating cultures.

Germin and Germin-Like Proteins from Higher Plants

Wheat germin (which is an OXO), is the best characterized
of all the cupin proteins in terms of its biochemistry, function,
and patterns of expression (45); it is therefore particularly
relevant to consider these various features in some detail. The
first evidence for such an enzyme that converts oxalic acid and
dioxygen to carbon dioxide and hydrogen peroxide came from
studies of powdered wheat grains in 1912 (320), although it was
more than 80 years later that the identity and sequence of this
enzyme were confirmed (173). In the meantime, there had
been two parallel and unrelated types of research concerning
this particular protein. The first of these concerned an impor-
tant medical application of considerable commercial signifi-
cance, namely, the use of barley OXO (98% identical to wheat
germin) in kits to assay levels of oxalate in blood plasma and
urine. Some of these kits (e.g., the Sigma kit) utilize an enzyme
isolated from barley roots, and although they are quick and
easy to use, there is a continuous effort to improve the accuracy
and efficiency of the assay (175, 191, 228). Such efforts will
benefit from recently obtained data regarding fundamental
biochemical and structural analysis of the barley enzyme itself
(161, 162, 238, 310) and from the finding (173) that the ex-
tremely well characterized wheat germin is also an OXO. This
discovery was the culmination of the second important re-
search track, one which started in the early 1980s, during which
the GF-2.8 germin (gi|121129) was found to be an apoplastic,
multimeric (310), glycosylated (135) enzyme with extreme re-
sistance to heat and to chemical degradation by protease or
hydrogen peroxide. These unusual properties have recently
been explained by the realization that wheat germin and its
relatives from barley and other cereals (206) are members of
the cupin family and that their resistance to extremes of envi-
ronment is likely to be a function of their structural similarity
to other desiccation-tolerant proteins including 7S and 11S
seed storage proteins; the resistance of the protein to H_2O_2 is
of course linked to its enzymatic generation of this compound.

Germin-like proteins (GLPs) have a maximum ca. 90% se-
quence identity (e.g., gi|1772596) to wheat germin, although
the average level of identity is closer to 50%. There is almost
complete identity in the conserved cupin core, in which the
intermotif spacing is 20 to 23 aa. Since the discovery of the first
GLP in a higher plant (127), there has been a rapid expansion
in the number of gene sequences identified, such that the latest
estimates give a total of 21 sequences in *Arabidopsis thaliana*,
the best-characterized plant genome to date (46; J. M. Dun-
well, unpublished data). However, no function has yet been
assigned to any of these sequences, with the single exception of
a *Pinus caribaea* GLP, which does have OXO activity (212). In
addition to the identification of GLP genes in analyses of
various plant genomes, expression of certain GLPs in plants,
including liverworts (gi|4718551) and mosses (gi|6042701,
gi|6102532), is associated with a range of specific developmen-
tal states but more particularly with specific biotic and abiotic
stresses, as detailed below.

**Germin-like proteins are expressed at specific developmen-
tal stages in plants.** Various studies have identified GLPs
during specific stages of plant development.

(i) Floral induction. Interesting evidence for the develop-
mental induction of GLPs in higher plants has come from
studies of floral induction; for example, a specific GLP tran-
script was found to show a circadian pattern of expression in
the long-day plant *Sinapis alba* (113) and its relative *A. thaliana*
(273). Similar results were obtained in the short-day plant
Pharbitis nil (218), where a GLP mRNA was detected specifi-
cally in the cotyledon and leaf. In a related study, the level of
a GLP in *Raphanus sativus* was found to be lower in young
flower buds than in leaf and root material (207), and a similar
GLP (gi|6090829) has recently been isolated from nectar of
Nicotiana plumbaginifolia (46a).

(ii) Fruit ripening. Studies of ripening fruit of mandarin
(118) (gi|1669031), strawberry (Dunwell, unpublished), and
apple (gi|3088119) have all reported finding GLP sequences.

(iii) Somatic and zygotic embryogenesis. Following initial
studies which identified several GLPs in embryogenic cultures
of Caribbean pine (*Pinus caribaea* Morelet var. *hondurensis*)
(59), a full-length GLP (gi|2745848) expressed in both somatic
and zygotic embryos was reported recently (212). Similarly,
GLP sequences have been found to be associated with somatic
embryos of Monterey pine (*Pinus radiata*) (gi|2935521), a sus-
pension culture of potato (gi|3171251), and a cell culture of
lupin (309).

(iv) Seed development. In a study (180) of proteins known to
provoke severe allergic reactions (part of the celery-birch-mug-
wort-spice syndrome), it was shown that the N-terminal se-
quence of the 28-kDa allergenic protein extracted from pep-
percorns of *Piper nigrum* has a high level of similarity (E value
4e-05) to a GLP (gi|2801803) from rice. This observation may
be linked to the fact that the well-characterized major peanut
allergen Ara h1 is a vicilin-like protein (258).

(v) Wood development. Recent studies (6) on a cDNA li-
brary produced from immature xylem from differentiating
wood in loblolly pine (*Pinus taeda* L.) identified a sequence
(gi|3365535) encoding a GLP similar (E value 3e-15) to an
Arabidopsis GLP (gi|1755152) and the *Physarum* spherulin
(gi|1052776). It is relevant that the largest group of sequences
with known function from this study were those associated with
cell wall formation and the lignin biosynthetic pathway, an
unsurprising conclusion in view of the fact that pine xylem is
characterized by massive cell walls. Similar studies (275) on

developing xylem elements of poplar (*Populus balsamifera* subsp. *trichocarpa*) also revealed two GLP sequences (gi|3857819 and gi|3858018).

Germin-like proteins are linked to specific plant-microbe responses. Evidence of a role for GLPs in the relationship between plants and microbes has come from studies of nodulation in legumes, as well as from investigations of specific pathogen responses in cereals.

(i) Nodulation in legumes. The first evidence for the occurrence of a GLP in a legume species came from a study of the mechanism of attachment of *Rhizobium* (and probably *Agrobacterium*) bacteria to the walls of plant cells, although this was not recognized as being so in the publication in question (284). The initial step in this non-host-specific attachment process involves rhicadhesin, a calcium-dependent (265) bacterial surface protein of about 14 kDa (264, 266, 285). Using an assay based on the suppression of rhicadhesin activity, a putative plant receptor molecule for this protein was purified from cell walls of pea roots (284). The N-terminal 29 aa of this protein were determined to be ADADALQDLC(?)VADYAS VILVNGFASK(O)(P/Q)LI. Although the authors of this study found no homology to known proteins, this sequence is very similar (69% identity; E value 0.006) to an *Arabidopsis* GLP (gi|1934730). Of particular relevance to the discussion elsewhere in this review is the observation that the receptor molecule was most easily removed from the cell wall with an aqueous solution of oxalate and oxalic acid. This finding suggests that the protein requires calcium for its anchoring, function, or stability and adds to the circumstantial evidence linking oxalate to the level of calcium in the cell wall and the consequent functional control of other proteins in that environment.

In addition to this evidence for the existence of a GLP related to bacterial attachment to the wall of legume root tips, it is known that oxalate itself is found at the very high level of 70 mM in faba bean (*Vicia faba*) nodules (294). Application of water stress to such nodules increases the level of bacteroid OXO fourfold and reduces the level of oxalic acid by 55% (295). It is suggested that the oxalate found in this location could act as a complementary substrate for bacteroids and as a means of slowing the decline in nitrogen fixation induced by water-restricted conditions.

(ii) Pathogen responses in plants. Plants defend themselves against pathogen attack by utilizing a variety of mechanisms that include the production of specific antimicrobial compounds, the cross-linking of lignin and proteins in the cell wall, the synthesis of cell wall-strengthening carbohydrate polymers, and hypersensitive cell death. Although a role in pathogen response was among the earliest of functions suggested for germin (170, 174), such a connection was not established until the identification of germin as an OXO, together with other studies on the interactions of powdery mildew, *Blumeria* (syn. *Erisyphe*) *graminis*, with leaves of barley (62, 63, 303, 322) and wheat (129). Subsequently, it has been shown that a specific pathogen-response OXO transcript is found in the wall of barley mesophyll cells 6 h after inoculation with mildew; the enzyme accumulates after 15 to 24 h (324). Additionally, a related sequence has been isolated from barley which shows papilla-mediated resistance to this disease (303). This particular transcript peaks at about 18 to 24 h after infection, specifically in the epidermal cells. Analysis shows that this temporal and spatial pattern f expression closely follows the formation of papillae, appositions formed on the inner surface of the epidermal wall and thought to be composed of proteins, polyphenols, callose, silicon, and guanidine-containing compounds. Such a composition is reminiscent of the complex spherule and capsule walls referred to above. It has been suggested that the H_2O_2 produced by the OXO members of this family may act as a messenger for activation of other defense genes in the same cell or in neighboring epidermal or mesophyll cells. It is also relevant to note the tenacious association between wheat germin and the arabinose-rich hemicelluloses (arabinoxylans or arabinogalactans) of cereal walls (135).

There is increasing evidence that there are common links between the transduction pathways for the detection of and response to biotic and abiotic stresses and that active oxygen species are involved in the plant-environment interaction (290, 308). In particular, the role of H_2O_2 in the generation of hydroxyl radicals (OH) has been proposed (84). In this context it may also be relevant to consider the potential role of the crystal idioblasts, specialized cells that contain crystals of calcium oxalate and occur throughout the leaves of many plants. It has been demonstrated (58) that certain pathogenesis-related proteins accumulate within these cells, and of course the supply of oxalate in these cells would provide a source of H_2O_2 if adequate levels of OXO were present.

Recently, the first circumstantial evidence linking a GLP to a pathogen response in a dicotyledonous species was reported (B. Fristensky, unpublished data); the EST sequence gi|4090021, found during a study of gene expression in leaves of *Brassica napus* infiltrated with pycnidiospores of *Leptosphaeria maculans* PG2, encodes a protein identical (with one frameshift) to the GLP1 gi|914911.

Germin-like proteins are induced by abiotic stress in plants. The first evidence for induction of GLP expression by abiotic stress was provided by a study of salt stress in barley roots (126, 128). Related results were subsequently obtained from the common ice plant *Mesembryanthemum crystallinum*, a facultative halophyte and a model (37) for the induction of Crassulacean acid metabolism during water stress and treatment with high levels of salt. It was found (1) that the oxalate content of the leaf bladder cells increased from <1 mM to 106 mM as salt levels were increased from 1 to 5 mM. These results may be related to the modulation of a GLP mRNA found during transcript analysis in this species (10, 204) and to the more recent identification of other similar ESTs (e.g., gi|3325551 and gi|4996622) in salt-treated plants. The link between oxalate metabolism and GLP induction is considered in detail below.

Among the most interesting of the cupin proteins related to abiotic stress is BspA (for "boiling-stable protein"), a 66-kDa protein highly expressed in cultured shoots of aspen (*Populus tremula*) exposed to water stress (222). This protein is also induced by abscisic acid application and by osmotic and cold stresses. In a recent study of greenhouse-grown plants (223) a lower level of expression of BspA was found in *Populus tomentosa* than in *Populus popularis*, a species more tolerant of water stress. It has been suggested that BspA contributes to membrane stability, a feature of considerable significance in relation to stress responses. Other abiotic stresses which recently have been shown to induce GLPs include manganese deficiency in tomato roots (gi|2979494; gene *Mdip1*), aluminum treatment in wheat (gene *war13.2*) (108), heat treatment in barley (298), and submergence in rice (gi|2952338, gi|3201969; see also tomato EST gi|28973890 and gi|5827572 from *Botrytis*). The most comprehensive of these studies is that utilizing a promoter-glucuronidase (GUS) fusion (27, 28) and showing induction of the wheat germin promoter in transgenic tobacco treated with salt, heavy metals, aluminum and plant growth regulators, specifically auxin and gibberellin.

Auxin-Binding Proteins

in-binding proteins (ABPs) (intermotif spacing of 24 aa) meric, glycosylated plant proteins encoded by a small family in each species. They are thought to act as a or for the auxin indole-3-acetic acid (141, 142, 300) and y to mediate a wide range of physiological responses ing a reduction in cytoplasmic pH in certain cells (93). sis of the gene structure reveals a four-intron/five-exon gement, with the central, third exon encoding the region includes the peptide responsible for binding the carboxcid group of indole-3-acetic acid. This motif, known as (41) or D16 (300), is now thought to be equivalent to the rved motif 1 in the cupin notation (69), a finding supd by observations on two similar proteins isolated from apices of peach (*Prunus persica* L. cv. Akatsuki) (217). e latter proteins have been designated ABP 19 916807) and ABP 20 (gi|1916809) on the basis of their y to bind auxin, albeit at low affinity (217). Recent analysis eir sequences shows a greater level of similarity to the s (the closest neighbor [E value 3e-78] is GLP3 1755164] from *A. thaliana*) than to any of the functionally er characterized ABPs.

Epimerases

another group of cupin enzymes involved in the synthesis of terial and archaeal cell wall components are the epimees, such as dTDP-4-dehydrorhamnose 3,5-epimerase (also own as dTDP-L-rhamnose synthase) (EC 5.1.3.13), which verts dTDP-4-keto-6-deoxy-D-glucose into dTDP-4-keto-6-oxy-L-mannose. These enzymes are about 185 aa in length d contain the two-motif cupin signature usually separated by listance of 28 residues; both motifs contain a single globally nserved histidine residue. They are encoded by *rfbC* (or uivalent), part of the *rfb* gene cluster (160, 189, 193, 205, 6). Most *rfb* operons start with an *rfbABCD* cluster, which is sponsible for the synthesis of TDP-rhamnose (184); this clus- r is followed by *rfbIFGH* in organisms that produce 3,6-deoxyhexoses.

These epimerases are located in the periplasm, and it is elevant to the theme of this review to note that periplasmic rotcins are, as a rule, folded into stable, protease-resistant onformations, consistent with the digestive nature of this ompartment (70).

Many of these capsular polysaccharides have potential ecoomic importance as aqueous rheological control agents for liverse industrial and food applications. Such compounds inlude xanthan gum (*Xanthomonas campestris*) (22), and the phingans (e.g., gellan, welan, and rhamsan) produced by spees of *Sphingomonas* (314). It has been proposed that the various sphingans be thought of as defensive in nature, similar to the protective capsules (224, 249, 277, 297) of many invasive pathogenic bacteria (e.g. alginate).

MULTIDOMAIN PROTEINS WITH A SINGLE CUPIN DOMAIN

In the multidomain proteins with a single cupin domain, the conserved cupin element does not lie at the core of the protein but instead represents a single domain in a complex multidomain organisation. The most notable group of proteins in this category consists of a subset of the AraC bacterial transcription factors.

AraC-Type Transcription Factors

Of all the bacterial transcriptional regulators, possibly the best characterized are the members of the AraC/XylS family (88). This family, named after its first member, AraC (a regulator of the arabinose pathway in *E. coli*), contains more than 100 members, which can be subdivided into various classes on a functional basis. These functions are associated primarily with carbon metabolism, stress responses, and pathogenesis, with the former category including factors that control the degradation of arabinose (AraC), cellobiose (CelD/ChbR), melibiose (MelR), raffinose (RafR), rhamnose (RhaR), and xylose (XylR).

Sequence analysis shows most members of this family to be 250 to 300 residues in length, comprising a conserved C-terminal of about 100 aa which binds DNA, and a nonconserved N-terminal domain which binds the effector molecule (44). There is much more information available on the DNA binding component, although the specific details of the N-terminal section (particularly of the AraC protein) are more relevant to the present review. This regulator has been subject to detailed structural (269, 270) and molecular (250, 255) analysis over several years. In summary, the N-terminal section comprises an arabinose-binding, eight-stranded β-barrel, which is joined to the DNA-binding domain via a linker region; the barrel-shaped section is also responsible for the dimerization of the molecule, a factor which determines its 3D shape and therefore its ability to bend the associated DNA strand. Close analysis of the sequence (90) and structure (Dunwell, unpublished) of this barrel-shaped element reveals a previously undetected similarity to the conserved β-barrel core of the cupin proteins (Fig. 2). Of this related subgroup of regulators involved in sugar degradation, that showing the closest sequence similarity to GLPs and other cupins is CelD (221). This protein was named on the basis of its presumed involvement in the utilization of cellobiose, although recent studies (150) have shown that the real function is as a regulator in the catabolism of the disaccharide chitobiose; on that basis, its gene has been renamed *chbR*, part of the *chb* (*N,N*-diacetylchitobiose) operon. The significance of this reassignment is that it further supports a functional link both to the other bacterial enzymes concerned with sugar metabolism (e.g., PMIs and epimerases) and to the higher-plant cupins, particularly the sucrose-binding proteins (detailed below). In this context, there is an additional circumstantial link between chitobiose and cupins, in that vicilins from cowpea (*Vigna unguiculata*) are known to bind chitin (248), and it has been suggested that the vicilin-induced inhibition of yeast cell growth is due to binding of the protein to the chitin component of the cell walls (96, 97).

TWO-DOMAIN BICUPINS

The first two-domain proteins recognized to be members of the cupin superfamily were the seed storage proteins (20); these are discussed below, particularly with reference to the structural analysis of cupins. More recently, several microbial proteins from archaea, bacteria, and fungi have been shown to have a two-domain cupin composition (64, 69), and this information has provided a new insight into the possible ancestral origin of the seed proteins. To distinguish the various subclasses of two-domain cupin, sequences are described in terms of their intermotif spacing and in terms of whether this spacing is the same (homo-bicupins) or different (hetero-bicupins) in the two domains.

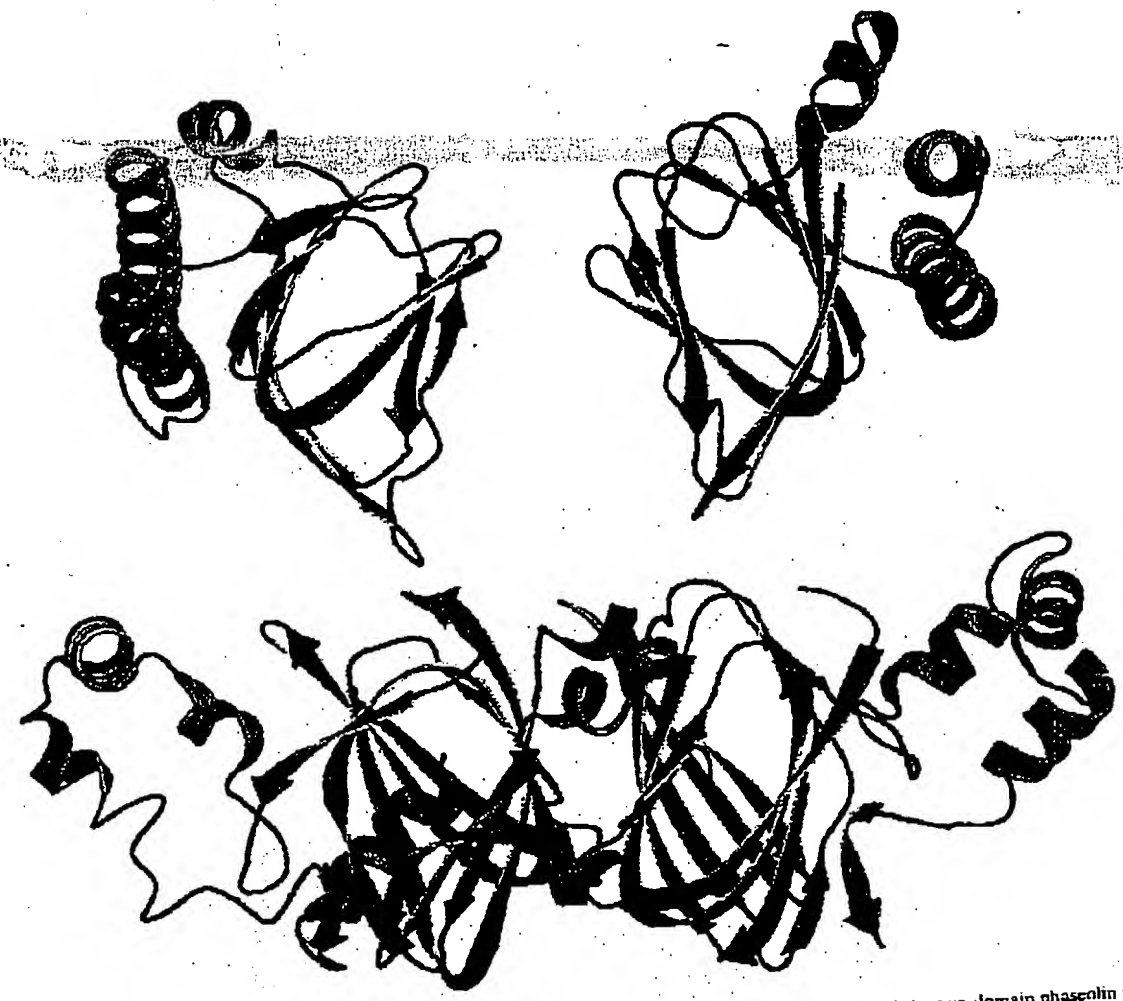


FIG. 2. Comparative structures of two orientations of the arabinose-binding domain of the AraC protein (above) and the two-domain phaseolin storage protein (below), showing the similar β-barrel element in the center of each domain, with associated α-helixes. The apparent gap in the E/F loop in phaseolin is due to the lack of resolution of the 3D structure at that point (177).

Gentisate 1,2-Dioxygenase and 1-Hydroxy-2-Naphthoate Dioxygenase

Identification of the two-domain composition of gentisate 1,2-dioxygenase (GDO) and 1-hydroxy-2-naphthoate dioxygenase (HNDO) is a novel finding made during the preparation of this review. The two enzymes are involved in the degradation of a range of related aromatic compounds, with the former enzyme, GDO (EC 1.13.11.4), catalyzing the oxygenolytic cleavage (between carbons 1 and 2) of gentisate (2,5-dihydroxybenzoate) to form maleylpyruvate, a compound that can be converted to central metabolites of the Krebs cycle either by cleavage to pyruvate and maleate or by isomerization to fumarylpyruvate and subsequent cleavage to fumarate and pyruvate. GDOs have been purified and characterized in many gram-positive and gram-negative bacteria (*Klebsiella pneumoniae* [143, 281], *Moraxella osloensis* [50], *Sphingomonas* [305], *and Actinomycetales* [109]), with possibly the best characterized such genes being those from species of *Pseudomonas* (110). For example, a GDO encoded by *nagI* (gi|3406827) has recently been identified in *P. aeruginosa* strain U2 (86) and a very similar polypeptide (E value 3e-45) is encoded by nucleotides 5549669 to 5548674 of a contig (gnl|PAGP_287|Paeruginosa_Contig54) from *Pseudomonas* strain PAO1. Another very similar sequence (gi|3293534) (Fig. 1) has also recently been found in *Haloferax* sp. strain D1227, an extreme halophile isolated from soil contaminated with highly saline oil brine and the only known aerobic archaeon able to utilize aromatic compounds as its sole carbon sources (85).

The only previous comment on the sequence similarity of these two types of dioxygenase was that made by Werwath et al. (305), who cloned the GDO gene *gtdA* (gi|3550667) from *Sphingomonas* sp. strain RW5 and showed that its product had a low similarity to the HNDO (EC 1.13.11.38) (gi|3288681) encoded by the *phdI* gene of the phenathrene-degrading *Nocardioides* sp. strain Kp7 (134). This latter enzyme catalyzes the cleavage of 1-hydroxy-2-naphthoate to *trans*-2′-caboxybenzalpyruvate, a ring cleavage between the carboxylated and hydroxylated carbons analogous to that effected by GDO.

Both classes of enzyme described in this section have a multimeric structure; GDO has an apparent subunit molecular mass of 38 to 39 kDa and is claimed to have either a tetrameric (85, 281, 305) or hexameric (151) composition, whereas HNDO has a molecular mass of 45 kDa and is considered to be hexameric (134). Like most other dioxygenases of the extradiol class (those that cleave an aromatic ring adjacent to two vicinal hydroxyl groups), both GDO and

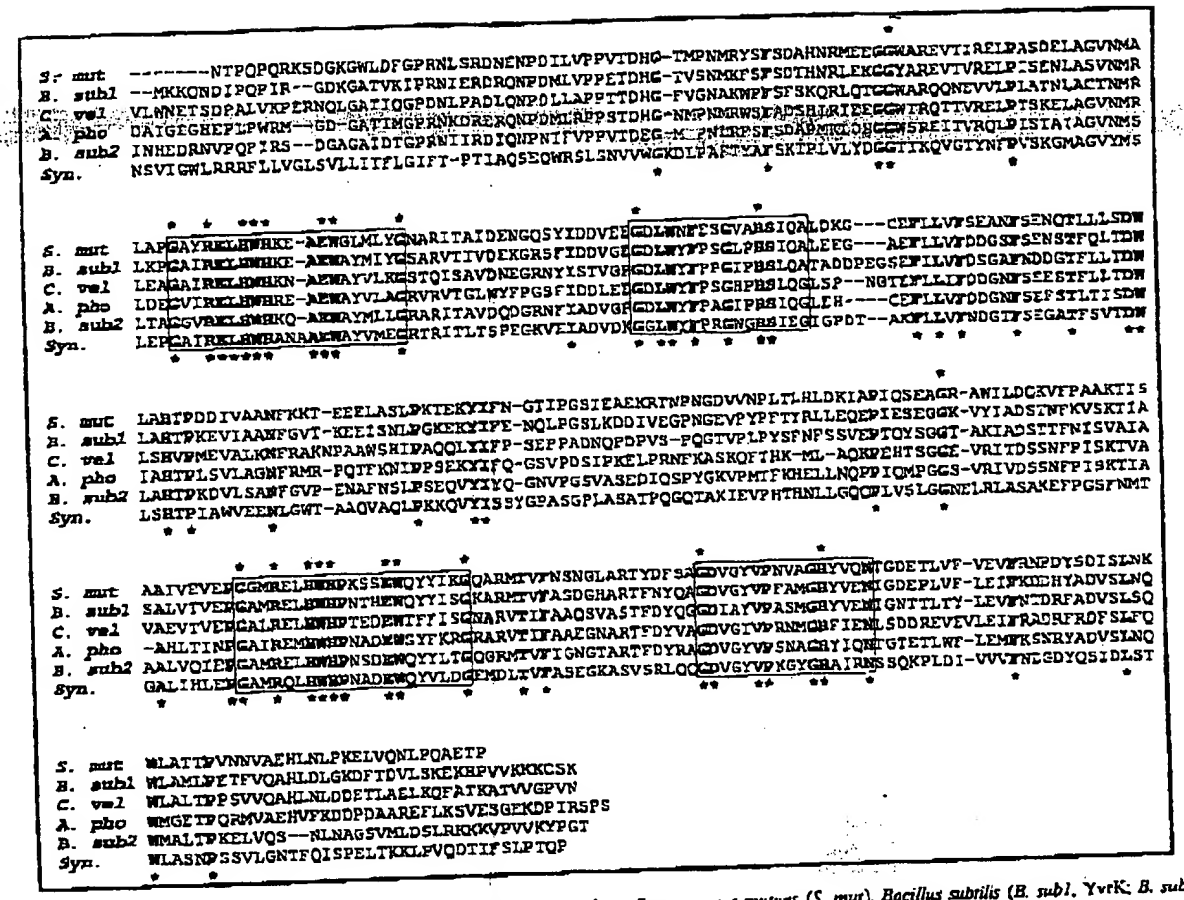


FIG. 3. Alignment of the six 20+20 bicupin proteins (presumed OXDCs) from *Streptococcus mutans* (*S. mut*), *Bacillus subtilis* (*B. sub1*, YvrK; *B. sub2*, YoaN), *Collybia velutipes* (*C. vel*), *Aspergillus phoenices* (*A. pho*), and *Synechocystis* (*Syn.*) (see the text for details), showing the positions of the two conserved motifs (boxed) within each of the two domains. Residues conserved in all sequences are indicated with asterisks below the alignment; residues also conserved between the two domains are indicated with asterisks above the alignment. The *A. phoenices* sequence (Scelonge and Bidney, PCT patent application WO 98/42827) has been amended by insertion of an additional nucleotide at residue 344 to correct a presumed frameshift error introduced during the sequencing of this gene.

HNDO contain 1 mol of Fe^{2+} per mol of subunit (those from *Arthrobacter globiformis* and *Bacillus brevis* contain manganese, although they utilize the same coordinating residues). These features, namely, a tetrameric or hexameric composition and the presence of a transition metal in the active site, are shared with other cupin proteins described in this review, such as barley OXO, which is now known to contain manganese (238, 239; S. Bornemann, personal communication).

Oxalate Decarboxylases

Among the many oxalate-degrading enzymes isolated from fungi, possibly the best characterized is that from the wood-rotting fungus *Collybia velutipes*. This particular homo-bicupin enzyme (intermotif spacing of 20 aa in each domain) degrades oxalate to formate and carbon dioxide and appears not to have any requirement for cofactors. It was therefore selected for use in strategies to reduce the levels of endogenous oxalate in plants (198, 199). The enzyme itself has an acidic pI, is stable over a wide pH range, is moderately thermostable, and has a molecular mass of 560 kDa as estimated by gel filtration and a subunit mass of 64 kDa before and 55 kDa after treatment with endo-β-*N*-acetylglucosaminidase, thus suggesting a glycosylated status (198). The sequence of the *C. velutipes* enzyme has been published as gi|1604990 (52), and recently the sequence of a similar enzyme from *Aspergillus phoenices* was reported (C. J. Scelonge and D. L. Bidney, 1 October 1998, PCT patent application WO 98/42827). Presumed homologues of these sequences have also been identified (Dunwell, unpublished) (see below) in the bacterial species *B. subtilis* and *Streptococcus mutans* (encoded by nucleotides 555 to 1676 from contig 1009) (Fig. 3).

Sucrose-Binding Proteins

Among the two-domain relatives of the seed storage proteins is a sucrose-binding protein (SBP) (gi|548900 and gi|2765097) found at low abundance in the plasma membrane of cotyledons, leaves, and mature phloem of legumes (103); a similar sequence (gi|2148163) from the cycad *Zamia furfuracea* is known (40). Recent comparison (219) of the soybean SBP sequence with that of vicilin has shown that the N-terminal domain of SBP contains 12 of the 13 residues conserved across the whole vicilin family, with the C-terminal domain having 10 of the 12 conserved residues.

Although the overall tertiary structure of SBP can be predicted by comparison to phaseolin, it is also possible that analysis of the disaccharide-binding domain of CelD/ChbR (see "AraC-type transcription factors" above) would provide further information on the specific ligands in the binding site.

TABLE 1. Summary of cryptic sequences encoding cupin proteins

Species	GenBank gi identifier (reference)	Nucleotide position	Closest neighbor, possible function
Archaea			
Desulfurococcus strain SY	1545808 (316)	104–448	gi|3256943 from *Pyrococcus horikoshii*, probable PMI
Thermococcus strain KS-8	4001717	1–349	As above
Eubacteria			
Aquifex aeolicus	2983162 (53)	7591–7914	15–164 gn1|TIGR|BTMBX89F *Thermotoga maritima*
Aeromonas caviae	4204207	232–23 (negative strand)	2632508 from *B. subtilis* AraC?
Mycobacterium leprae	1377767	19060–19155	gi|2984230 from *A. aeolicus*
Alicyclobacillus acidocaldarius	39300 (157)	2196–1755 (negative strand)	gi|1907078 human pirin (304)
Streptomyces lividans	48953 (234)	171–1 (negative strand)	As above
Bacillus stearothermophilus	560029 (216)	338–1 (negative strand)	146366–146533 contig 230 from *P. aeruginosa*
Desulfovibrio desulfuricans	49285 (278)	238–726	16083–115762 contig 272 from *Streptococcus pyogenes*
Pseudomonas lemoignei	531465	1618–1286 (negative strand)	82–405 contig 854 from *Bordetella pertussis*
Morganella morganii	508518 (60)	1666–1184 (negative strand)	gi|2632720 *ydaE* from *Bacillus subtilis*; sugar alcohol metabolism?
Corynebacterium glutamicum	2342561 (137)	3686–3357 (negative strand)	Pepl protein gi|1572541 from *E. acrogenes*
Synechocystis	287460 (178)	3–208 (add G at 104)	C terminus of gi|1652486, a PMI, same species
Azorhizobium caulinodans	763059 (92)	1–201	gi|1772621 from *Erwinia chrysanthemi*
Mycobacterium genavense	2558999	947–1519 (ATG at 960)	As above
Eukaryota			
Dictyostelium discoideum	1177288 (241)	302–3 (negative strand)	60364–60077 contig 229 from *P. aeruginosa*
Arabidopsis thaliana	987518	1195–1873 (add G at 1288)	gi|1755168 and gi|16847 from *Arabidopsis*

Seed Storage Proteins

During the development of plant seeds there is a massive accumulation of nitrogen and carbon reserves in the form of proteins that can withstand desiccation and be used as a source of energy for the germinating embryo. In legumes, the globulin type of storage proteins can be divided into two forms, the legumins and the vicilins. The former are usually found as hexameric complexes (sedimentation coefficient, 11S), with each subunit derived from a precursor complex consisting of two domains, an N-terminal acidic α chain and a C-terminal basic β chain, which remain associated following proteolytic processing. The latter proteins occur as 7S trimers, with each subunit being a 50- to 70-kDa polypeptide that is subject to variable levels of processing. Examination of Fig. 1 shows that most of the storage proteins either lack any of the conserved His residues or contain a single conserved His in motif 1. It is presumed that, as a consequence, they have no metal-binding ligands and therefore no enzymatic activity. There is, however, a massive accumulation of oxalate (maximum 24% [dry weight]) during early seed development in soybean (131) and presumably in other legumes, and it is tempting to speculate on the possibility that this compound acts as a substrate for a residual oxalate-degrading capacity provided by the storage proteins being produced at that period. Knowledge of the tertiary structure of the two storage proteins phascolin (177) and canavalin (155) and the finding of certain globally conserved residues (20) provided the basis for the generation of a homology model of wheat germin (90) and all subsequent predictions of cupin structures (Fig. 2).

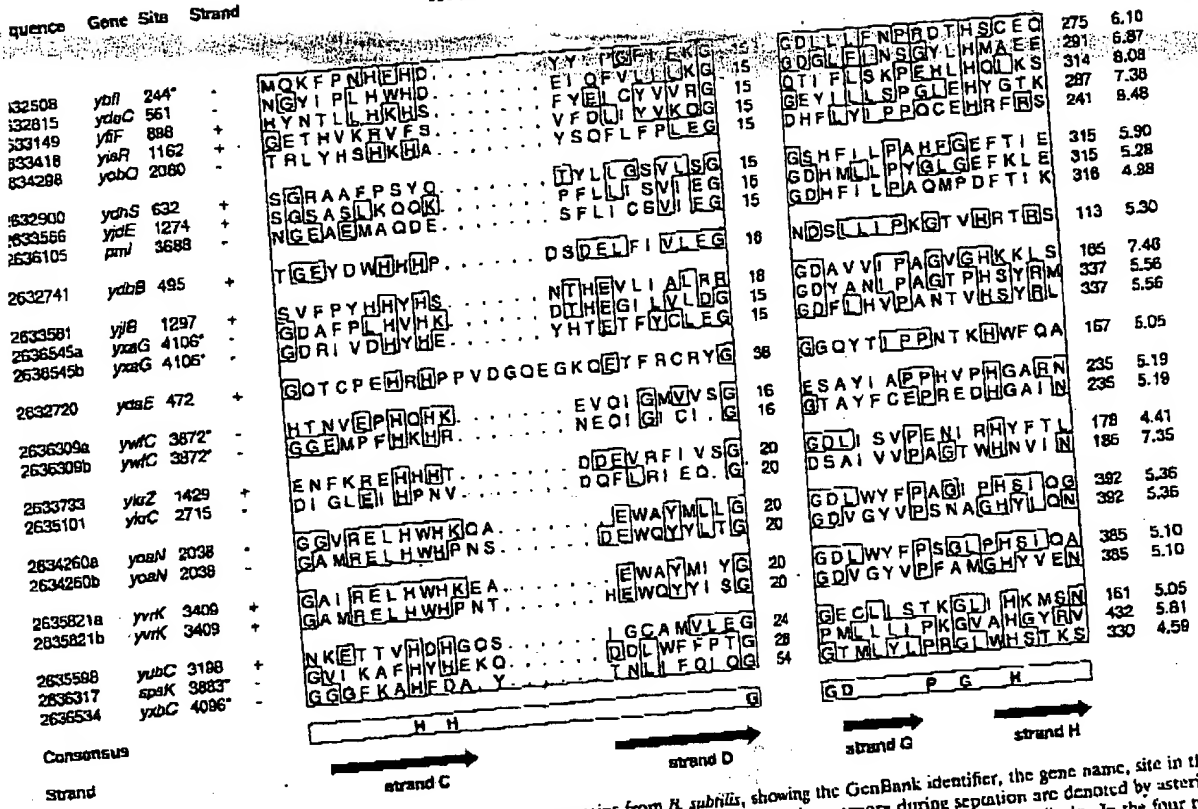
In addition to the well-known major storage proteins found in seeds and spores (261), other, less abundant proteins of this type have been the subject of detailed analysis. Among the best characterized is the major peanut allergen Ara h1, a member of the vicilin family (43, 54, 258) and the protein responsible for the majority of cases of fatal food-induced anaphylaxis. In a recent study (258), it has been shown using molecular modelling that the 23 linear immunoglobulin E-binding epitopes cluster into two main regions, thus providing a rational target for transgenic approaches (66, 67) to modify the allergenic residues. Like many other members of the cupin family described in this review, the Ara h1 protein has a very high level of stability; it survives intact in most food-processing methods and also resists digestion by the gastrointestinal tract or its in vitro equivalent (23). It has been suggested (258) that this stability may be due to its compact structure, which limits the possibility for protease digestion and also facilitates its passage across the small intestine. It is presumed that these biophysical characteristics are shared by the allergenic single-domain GLP recently identified in ground black pepper (180).

Bicupins of Unknown Function

As described above, there is now good evidence for a wide variety of bicupins from archaeal species (e.g., the GDO from *Haloferax* [85]), many bacteria including *B. subtilis* and *Streptococcus pyogenes* (the 15+15 bicupin encoded by contig 272) and several eukaryotes (e.g., seed storage proteins). With the exception of the two classes of dioxygenase and the OXDCs from *Collybia velutipes* and *Aspergillus phoenices*, no biochemical function has yet been assigned to the microbial bicupins. It would be of particular interest to investigate the activities of the four examples from *B. subtilis*, which now probably represents the best organism for the study of prokaryotic cupin diversity. Within the higher plants, there is also evidence for another previously unidentified class of bicupins (e.g., the *Arabidopsis thaliana* hypothetical gene gi\|2244827).

CRYPTIC SEQUENCES ENCODING CUPIN PROTEINS

In addition to the cupins described in the above section, there is a group of other related coding sequences (Table 1) (Dunwell, unpublished) not previously identified in the databases. These are either complete or partial ORFs, often found in apparently noncoding regions of other genes. These cryptic ORFs can be divided into various types, according to the rea-



IG. 4. Alignment of the conserved two-motif signature in the cupin proteins from *B. subtilis*, showing the GenBank identifier, the gene name, site in the genome kilobases from origin; sequences likely to be included in the section of the chromosome trapped in the prespore during septation are denoted by asterisks) (311). ing strand, details of the two motifs, total size of the protein, and its calculated pI. The sequences are subdivided on the basis of similarity. In the four two-domain ins (YxaG, YwfC, YoaN, and YvrK), the first and second domains are designated a and b, respectively.

on for the previous lack of identification. In one case, that om *Mycobacterium genavense*, it seems obvious that the in- orrect start codon was selected and thus a protein with no nown similarity was generated. In contrast, the nonannotated ORF (NORF) in *Aquifex aeolicus* was simply not identified by he algorithms used to find ORFs in such bacterial genomes 53). The occurrence of NORFs is well known from other complete genome or transcriptome studies such as that con- ducted on yeast, where serial analysis of gene expression tech- niques identified 160 NORFs (299). Presumably, the other examples identified in the present study were overlooked pre- viously simply because the ORF is in a reading frame different from that used by the gene which was the main subject of the specific study. In most cases, however, the analysis is also complicated by the inclusion of one or more frameshift errors in the sequence.

ANALYSIS OF CUPIN SEQUENCES IN *B. SUBTILIS*

Although the broad-ranging surveys described above are of considerable value in determining the overall occurrence of members of the cupin superfamily across various taxa, it was considered particularly important to conduct a detailed survey of a single prokaryotic genome in order to assess more accu- rately the spectrum of cupins encoded by such a genome. It was already known (64, 69) that archaeal genomes contain only a few (2 to 7) cupin genes, whereas the cyanobacterium *Synecho- cystis* has a complement of 18 cupins genes including one encoding a bicupin (65). Preliminary studies (Dunwell, unpub- lished) had suggested that *B. subtilis* was probably the most appropriate organism for this analysis since its genome en- coded a greater variety of plant-related cupins.

Overall Conservation of Cupin Motifs in Proteins Encoded by the *B. subtilis* Genome

Analysis of the genome of *B. subtilis*, using the methods described above, identified a total of 20 sequences that fulfil, at least in part, the characteristic two-motif cupin signature. The alignment of this conserved section is given in Fig. 4, which also shows the range of intermotif spacing (15 to 54 aa) as well as the overall protein size (113 to 432 aa). It can be seen that the sequences fall into several subgroups on the basis of their detailed similarity, with the great majority having the charac- teristic signature of three histidines (two in the first motif and one in the second), along with conserved proline and glycine residues in the second motif.

Particular reference must be made to YdaE (gi|2632720), which is most unusual in having an additional six residues between strands C and D within motif 1. It also has a compar- atively long intermotif distance.

Closest Neighbors and Possible Functions

Only two of the cupins in *B. subtilis* have designated names (PMI [phosphomannose isomerase] and SpsK [spore capsule synthesis K protein]); most of the sequences are so-called y genes (166), i.e., genes of unknown function that make up 70% of the total gene complement. The closest neighbor for each

TABLE 2. Analysis of the closest neighbors for each of the cupin sequences from *B. subtilis*[a]

Sequence (gi)	Closest neighbor		Length (aa)	Identity (%)	Similarity (%)	Gaps (%)	P	Function
	Species	(gi)						
2633149	*B. subtilis*	2633014	96	31	58		5e-11	AraC
2632815	*L. monocytogenes*	2745844	262	24	47	7	5e-17	AraC
2633418	*B. megaterium*	2764541	265	25	44	12	9e-14	AraC
2632508	*E. coli*	132526	248	23	41	8	8e-13	AraC
2634298	*P. leiognathi*	2495367	243	27	44	10	5e-17	AraC
2632900	*B. subtilis*	2636105	311	55	67		e-102	PMI
2633556	*B. subtilis*	2636105	316	56	70	<1	e-106	PMI
2636105	*B. subtilis*	2632900	311	57	69		e-107	PMI
2636545	*A. aeolicus*	2984227	61[c]	32	61		2e-04	
2632720	*M. morganii*	508518[b]	155	41	55	14	e-18	
2632741	*S. meliloti*	Unfin.[c]	88[d]	42	61	7	2e-12	?PKS
2636309	*E. faecalis*	Unfin.[c]	86[d]	26	44	2	0.36	
2633581	*D. radiodurans*	Unfin.[c]	136	37	57	2	5e-21	
2633733	*A. aeolicus*	2984230	175	33	57	8	3e-23	
2635101	*P. persica*	1916809	128	22	38	14	0.11	GLP
2634260	*B. subtilis*	2635821	379	58	75		e-134	
2635821	*B. subtilis*	2634260	379	58	75		e-134	
2635598	*C. albicans*	Unfin.[c]	101	32	50	5	2e-07	?CDO
2636317	*A. actinomyc.*	Unfin.[c]	270	41	62	4	9e-66	dTDP-DR[e]
2636534	*P. aeruginosa*	Unfin.[c]	236	23	40	0	2e-10	

[a] Estimated by use of the gapped BlastP program.
[b] This sequence was identified by a TBlastN search and encodes a previously unidentified polypeptide (see the text for details).
[c] These sequences were identified by use of a TblastN search of the unfinished microbial genomes database, see text for details.
[d] These relatively short regions span the conserved two-motif section of the protein.
[e] dTDP-4-dehydrorhamnose reductase.

protein sequence, as estimated by a BlastP analysis, is given in Table 2. In terms of function, it can be seen that the sequences can be divided into various subgroups that include five AraC-type transcription factors, three PMIs, and a cysteine dioxygenase. However, an obvious problem inherent in this type of comparison based on the total sequence is that it takes no account of the occurrence of multidomain proteins. For example, analysis of sections of the SpsK protein suggests that it probably represents a bifunctional enzyme similar to one from *Actinobacillus actinomycetemcomitans*, with an N-terminal domain presumed to have dTDP–4-dehydrorhamnose reductase activity (cf gi|2650312 from *Archaeglobus fulgidus*) and a C-terminal domain (containing the cupin element) with dTDP–4-dehydrorhamnose 3,5-epimerase activity (c.f. gi|2622921 from *Methanobacterium thermoautotrophicum*).

The unusual protein YdaE is most closely related to a previously unidentified protein from *Morganella morganii*.

Additional confirmation of the different functional subgroups can be obtained by examination of the pI values given in Fig. 4. This shows that all the transcription factors have values between 6.10 to 8.48 whereas the other proteins (with the exception of YjlB and YrkC) are more acidic, with values between 4.41 and 5.90.

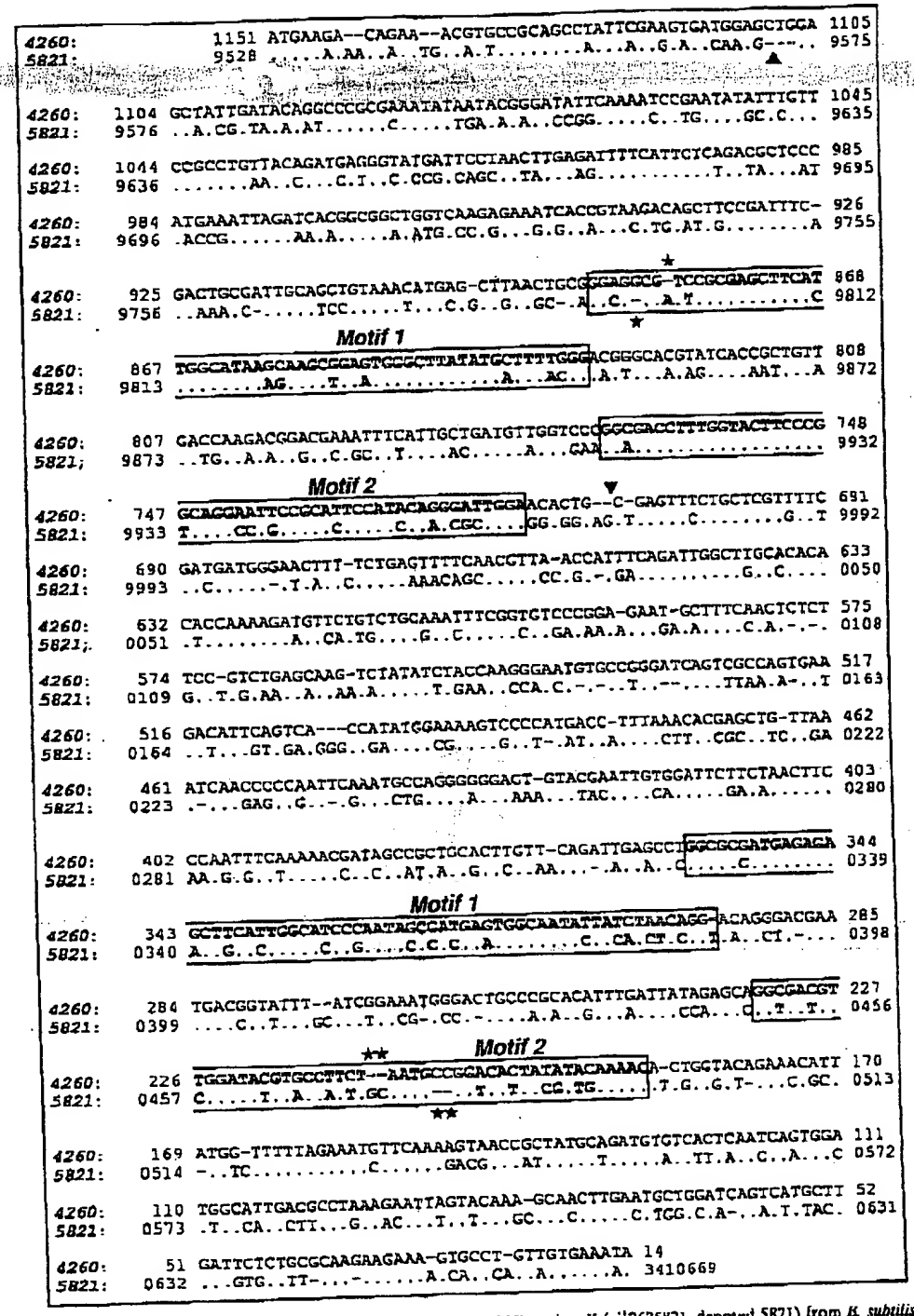
Domain Structure

There are 16 single-domain and 4 two-domain (bicupin) proteins encoded by the *B. subtilis* genome (Fig. 4). Bicupins are referred to below on the basis of their intermotif spacing (e.g., 15+15, 20+20). Of the former group of one-domain sequences, particular note should be made of the two examples that have a spacing of 20 residues, namely, YkrZ and YrkC. The former is most similar to a recently described sequence from the hyperthermophilic bacterium *Aquifex aeolicus* (53), whereas the second sequence is closer to a sequence from *Prunus persica*.

Probably the most interesting of the latter group of bicupins are the two sequences YoaN and YvrK, which have a very high level of similarity (E value 1e-130) to a sequence from *Streptococcus mutans* (contig 1009) and to the oxalate decarboxylases encoded by gi|1604990 from *Collybia velutipes*, a wood-rotting basidiomycete (198), and the related sequence from *Aspergillus phoenices* (Seelonge and Bidney, patent application). These fungal enzymes are related to the *Synechocystis* protein gi|1652630 (69), the only other 20+20 microbial bicupin identified to date. Detailed inspection of the six-sequence alignment provided in Fig. 3 reveals two main features. First, there are 64 (c. 16% of the total) globally conserved residues, mostly clustered within the two cupin motifs, which have the composition $GX_2RX_2HWHX_{3/4}EWX_5G$, and $GX_{10}HX_4$. Of these 64 residues, only 11 (ca. 3%), including the 3 histidines (90), also show conservation between the first and second domains. Second, the fungal OXDCs are more similar to the sequences from *B. subtilis* and *S. mutans* than they are to the *Synechocystis* protein.

Additional alignments of protein sequence (data not shown) suggest that the most likely single-domain progenitor of the two-domain 20+20 proteins is YkrZ and that this protein is slightly more similar to YvrK than to YoaN. The evolutionary time course of events is thus indicated to be (YkrZ) × 2 → YvrK → YoaN. Similarly, it is likely that YjlB (18 spacing) is the progenitor of its closest neighbor, the two-domain YxaG (15+15) sequence (Table 2), although this would imply that the increase in intermotif spacing from 15 to 18 residues in YjlB occurred after the duplication event. It is also noticeable from the alignments of single cupins with their putative bicupin derivatives that the single-domain sequences (e.g., YjlB) always show a higher degree of similarity to the C-terminal domain than to the N-terminal domain of the respective bicupin (e.g., YxaG).

If alignments are based on the DNA rather than the protein sequence, additional features can be observed (Fig. 5). For example, the doublet of bicupin genes (*yvrK* and *yoaN*) are very



. 5. Alignment of the two 20+20 bicupin genes *yoaN* (gi|2634260, denoted 4260) and *ywrK* (gi|2635821, denoted 5821) from *B. subtilis*, showing the positions two conserved motifs within each of the two domains. Similar nucleotides are shown as dots, and deletions are shown as dashes. The deletions marked with ks in motif 1 of the first domain (1-bp deletion) and motif 2 of the second domain (2-bp deletion) denote two examples of the compensatory system of deletions sertions that maintain the same reading frame for the two genes throughout the majority of the sequence. The deletions which produce a respective difference presence of a Gly residue (Fig. 3) are marked with vertical arrowheads. The statistical analysis is as follows: score = 1,706 (256.0 bits), expect = 2.2e-72, *P* = 2, identities = 756/1,151 (65%).

lar to each other (65% identity; E value 2.2e-72), although gene has a different pattern of insertions and deletions els). However, these differences in nucleotide sequence do disrupt the conserved two-motif regions; where there are indels within these motifs, they are equivalent in the two genes and do not alter the globally conserved residues.

In an earlier study (69) it was suggested that the two-domain OXDC proteins may represent direct progenitors of the two-

domain storage proteins. Recent phylogenetic evidence (260) now shows that the two duplication events occurred independently.

Physical Location of Cupin Genes within the *B. subtilis* Chromosome

The cupin sequences are arranged on both DNA strands, and although they are distributed throughout the chromosome (Fig. 4), there is a possible increase in the kilobase value as the complexity of the protein increases. It is also noticeable that the two members of the doublet (*yvrK* and *yoaN*) are on opposite strands and that all four of the two-domain sequences are located in the second half of the chromosome (i.e., above kb 2000).

SUMMARY OF GENOME ANALYSES OF *B. SUBTILIS* AND OTHER ORGANISMS

There are several important conclusions to be drawn from this study on *B. subtilis*. Most importantly, it has identified a previously unrecorded grouping of 20 cupin genes (0.5% of the total of 4,100) in the archetypal gram-positive species. This group of sequences provides evidence for two types of gene duplication having occurred during the evolution of the *B. subtilis* genome and/or the genome(s) of its progenitor(s). First, there has been duplication to increase the number of cupin genes. It is estimated (166) that *B. subtilis* has 568 (14%) of its 4,100 genes in the form of doublets and 273 (7%) in the form of triplets. In the present study, the most obvious example of a doublet is *yoaN* and *yvrK*, the genes encoding two-domain proteins closely related to the fungal OXDCs. Similarly, *pmi* and its two related sequences are members of a triplet, and the five genes encoding AraC-type transcription factors with identifiable cupin motifs are representatives of an even larger gene family (it is estimated that *B. subtilis* has a total of 11 members of this class of transcription factor).

Second, there is evidence of duplication or fusion to produce the genes encoding the two-domain bicupin proteins. Moreover, this process must have occurred on at least three occasions to produce YxaG (15+15), YwfC (16+16), and YvrK/YoaN (20+20). One would assume that the chronological order for these duplications was in the order 2 × 15, 2 × 16, and 2 × 20, although there is no certainty for this assumption. It is also unclear whether there is an extant progenitor for each of the bicupins or whether the immediate progenitor was replaced by its two-domain successor.

The other major conclusion to be drawn from this study is that in comparison with other microbial genomes sequenced to date, the *B. subtilis* genome is the closest to the eukaryotes in terms of both the overall number of cupins encoded and their specific domain composition. Of particular interest is the finding that one of the single-domain *B. subtilis* sequences (gi|2635101) has a GLP sequence from an angiosperm (Table 2) as its closest relative. It should be noted that this bacterial sequence, together with the similar sequence (gi|3258400) from the archaeon *Pyrococcus horikoshii* (Fig. 1), were not available in the previous study (69). They are thus the first 20-aa-spacing cupin sequences to have been identified in prokaryotes. Together with the very similar 19-aa proteins (gi|2984230 from *Aquifex aeolicus* and the polypeptide encoded by nucleotides 46806 t 47348 in contig 235 from *Pseudomonas aeruginosa*), they are of great significance to our understanding of the evolutionary origin of the increasingly important eukaryotic GLPs, which in plants are encoded by a gene family comprising many members. Such observations confirm that there has been large-scale multiplication of this type of cupin sequence during the course of evolution. It is known that GLPs occur in several species of gymnosperm, including *Pinus caribaea* (59, 212), *P. radiata*, and *P. taeda* (6). It is not known, however, whether this expansion in numbers, from two members in *B. subtilis* to a large family of genes in higher plants, occurred before or after the divergence of the angiosperms from the gymnosperms.

EVOLUTIONARY ASPECTS OF CUPIN COMPOSITION IN MICROBIAL GENOMES

Size of Cupin Gene Families in Prokaryotes and Eukaryotes

In general terms, it can be seen that the number of cupin genes varies from 2–7 in the archaeal genomes sequenced to date and in *Aquifex aeolicus* (often considered to be the most deeply branching bacterium) to 15–20 in *B. subtilis* and *Synechocystis* (65) to >40 in angiosperms and animals (Dunwell, unpublished). For example, in *Arabidopsis thaliana* there are about 20 GLPs, which themselves can be further subdivided on the basis of sequence into five to seven subclasses. Apart from the identification of OXO activity in some of the cereal germins and in a related protein in *Pinus caribaea* (212), there is no functional information about these particular eukaryotic proteins.

Interestingly, with the exception of the epimerases (e.g., gi|1666505 from *Leptospira interrogans*), cupins appear to be absent from spirochetes, as estimated by the analysis of the relatively small genomes of *Borrelia burgorferi* (genome size, 1.44 Mb) (79), *Treponema pallidum* (1.14 Mb) (80), *Rickettsia prowazekii* (the closest extant relative of the ancestor to mitochondria, with a genome size of 1.10 Mb) (9), and *Chlamydia trachomatis* (1.05 Mb) (274); for reference, the genome size of *B. subtilis* is 4.20 Mb. Presumably, this specific absence of cupins is part of the reductive evolution that leads to the general absence of any genes for cellular biosynthetic functions, including most aspects of cell wall synthesis, in these intracellular parasites.

It should also be noted that eukaryotic cupins are not restricted to fungi, yeasts, and plants. The latest estimate of their abundance in animals (68) suggests a widespread occurrence, with *Caenorhabditis elegans* having a total of at least 15 sequences (all single domain), including some (e.g., gi|3877049) which have a 20-aa spacing and are very similar to sequences from *A. thaliana* (gi|2739365), rice (gi|3201969), and *H. sapiens* (gi|2822126). In terms of function, some of these animal sequences, such as PMIs, cysteine dioxygenases, and epimerases, are related to microbial sequences described above, while others, such as the CENP-C centromeric proteins (282) and various zinc finger transcription factors, are specific to eukaryotes.

Do Cupin Families Arise from Gene Duplication or Genome Fusion?

In the discussion above, it has been assumed that the gene duplication events to generate doublets and higher-order multiples all took place in a single bacterial genome. However, there is a possibility that the present genome of *B. subtilis* is the result of one or more fusions of ancestral genomes (either in whole or in part) and that the doublets of genes represent the consequence of this fusion process (130). Large-scale analyses of protein sequences have already led to the suggestion that the evolution of the archaea included at least one major merger between ancestral cells from the bacterial lineage and the lineage leading to the eukaryotic nucleocytoplasm (158,

40, 242). In the same context, it has also been proposed that the eukaryotic nuclear genome is a chimera that has ed major contributions from both an archaeal species gram-negative species. Such a network of fusion events clearly lead to a complicated structure for many pro- tic genomes, with gene doublets resulting from either or internal duplication.

ical Location of Cupin Genes in the Bacterial Genome

important aspect of cupin evolution concerns the fact upin genes are distributed on both DNA strands of the *B. is* chromosome. This suggests that at some stage during ition, there was a duplication of genes from one strand to ther. If the specific example of the 20+20 bicupins is ined, for example, it can be seen that in *Synechocystis* is one such gene on the left strand whereas in *B. subtilis* two closely similar genes are transcribed by different ds (Fig. 4).

nalysis of the location of the cupin sequences within the *B. lis* genome (Fig. 4) shows a general progression, with the s encoding the shorter proteins being located closer to the in of replication than are the genes encoding the longer eins. Similarly, all the two-domain sequences are located e second half of the genome (from kb 2038 to 4106). There tle information on any factor(s) that might determine the er-order location of genes in this species, although it has n suggested (166) that the "grey hole" located at kb ~600 ht be related to the temporary chromosome partition ob- ed during the first stages of the sporulation, when a seg- t of about one-third of the chromosome enters the pres- e and remains the sole part of the chromosome in the spore for a significant transition period (311). In light of the eral link between cupin proteins and sporulation/desicca- (20, 69), it is possible that there is a functional reason for clustering of genes required during this stage of the life le.

There are two additional pieces of evidence for physical kage between related cupin genes and between such genes other functionally related genes. Most interestingly, it ms likely that the recently sequenced genome of *A. aeolicus*) represents the earliest known stage of bacterial cupin olution, a suggestion supported by the fact that the family *uificaceae* is the most deeply branching family within the cterial domain on the basis of phylogenetic analysis of 16S NA sequences. Specifically, *A. aeolicus* genome contains ree cupin genes. Between aq_528 and aq_529 there is a ptic cupin gene encoding a protein with a 15-aa spacing nilar to gi|2128971 from *Methanococcus jannaschii*, and in ldition there are two other closely linked cupin genes q_1975 and aq_1978). The first encodes a protein with a -aa spacing (gi|2984230), and the second encodes a protein ith a 15-aa spacing (gi|2984227); both have close relatives in . *subtilis* (Table 2). It seems most likely, therefore, that q_1975 was the product of duplication of aq_1978 and that it en diverged in sequence by the addition of 12 bp (encoding aa) to the intermotif region while retaining its close physical osition in the chromosome. In contrast, their two paralogs in . *subtilis* (*yjlB* and *ykrZ*) are located more than 200 kb apart, resumably as a consequence of recombination.

In this context, there is also an interesting relationship be- ween the *ydbB* gene and its adjacent sequences in the *B. ubtilis* genome. A TBlastN analysis of this gene shows that its losest relative from plants is the wheat protein germin, so named because of its high level of expression in germinating embryos (172). Adjacent to *ydbB* is *gsiB* (for "glucose starvation-inducible protein B") (gi|2632740), a sequence that has as its closest neighbours (E values 2e-18, and 2e-15, respectively) the plant Em protein (gi|1169515) from *A. thaliana* (91) and the Lea (late embryogenesis abundant) protein (gi|547819) from barley (75, 243). It is possible, therefore, that the two adjacent bacterial sequences have a common developmental link that began with a role in stress response, was retained throughout evolution, and is now associated with embryo development in higher plants.

Comparison of Single-Domain and Two-Domain Cupins

Detailed analysis of various pairwise alignments of the single-domain and two-domain cupins show a number of interesting features relating to the possible origin and evolution of these proteins. First, as described above, single-domain proteins are more similar to the C-terminal domain than to the N-terminal domain of their respective two-domain relatives. The reason for this disparity is unknown, but it is possible that there is less selection pressure to conserve the sequence of the N-terminal domain in a two-domain cupin. Another possible explanation for such variation is suggested by similar discoveries in the family of extradiol dioxygenases, where there are also single- and double-domain members (74). Among some of the latter enzymes, there is evidence that the two domains express different phylogenies, suggesting the possibility that these particular enzymes arose from recombination or even fusion of genes encoding different dioxygenases.

Cupins and the Comparative Structure of Microbial Cell Walls

The very close similarity between the two-domain 20+20 bicupin (gi|1652630) from the gram-negative cyanobacterium *Synechocystis* (65) and the related sequences from *B. subtilis* and *Streptococcus mutans* (Fig. 3) supports the observation (105) that the cyanobacteria constitute one of the deepest-branching clades within the gram-negative species and have a close affinity to the gram-positive species. Indeed, analysis quoted by Gupta (105) and based on sequences from glutamate dehydrogenase, phosphoribosyl formyl glycinamidine synthase, and some 16S rRNAs suggests a clade comprising the gram-positive bacteria, the cyanobacteria, and the archaea. Additional circumstantial evidence for a close link between some gram-positive bacteria and some archaea comes from an analysis of the archaeal cupins. Whereas only two such sequences have been identified in *Methanococcus jannaschii* (42), seven have been found in *Methanobacterium thermoautotrophicum* (Dunwell, unpublished). It may be that the additional cupins in *M. thermoautotrophicum* are linked to its particular cell wall characteristics, since it was noted by Gupta (105) that a number of archaea, including *Methanobacterium*, exhibited a thick and homogeneous cell wall that shows positive staining in the Gram reaction. Perhaps the *M. thermoautotrophicum* sequences with greatest relevance in the context of wall complexity are the two hypothetical proteins gi|2621649 (89 aa) and gi|2621650 (89 aa) that contain sequences with similarity to the first and second motifs of the *spsK* (for "spore capsule synthesis K" [gi|2636317]) protein from *B. subtilis*. Indeed, close analysis of these two adjacent archaeal ORFs suggests that they are more likely to be contiguous sections of a single coding region, particularly since addition of a single nucleotide (to correct a frameshift) produces a modified, longer ORF encoding a polypeptide most similar (E value 3e-19) to the C-terminal 151 aa of this *B. subtilis* protein, a probable dTDP-4-dehydrorhamnose 3,5-epimerase (cf gi|2622921). Such enzymes are involved in the rhamnose biosynthetic pathway

linked to the production of complex exopolysaccharides in cell walls.

In the latest discussion of prokaryotic classification, Gupta (106) has used differences in cell wall architecture to argue against the concept of considering the *Archaea* (307) to be a separate kingdom and in favour of linking the gram-positive bacteria and the archaea in a grouping of monoderm prokaryotes (surrounded by a single membrane) distinct from the diderm prokaryotes (i.e., all true gram-negative bacteria containing both an inner cytoplasmic membrane and an outer membrane). It is possible that the functional role of many cupins in cell wall synthesis will help in the resolution of this debate.

STRUCTURAL ASPECTS OF CUPINS

Despite the wealth of information on the primary sequence of cupins and their conserved core motifs, relatively little is known about their secondary, tertiary, and quaternary structure. As described above, the major advance in this area came from the discovery (20) of the global conservation of a small number of residues in the plant storage proteins and germins and the slime mold spherulins (29). This discovery enabled the 3D structure of the bean storage protein phaseolin (177) to be used to generate a homology model (90) of the wheat GF-2.8 germin, an OXO, and to predict various quaternary structures for the arrangement of subunits. Comprehensive physicochemical studies (197) of the monomer and oligomer had suggested a homopentameric assembly of subunits in native wheat germin, but subsequently X-ray diffraction studies of barley germin in the same laboratory (E. F. Pai and B. G. Lane, unpublished [but cited in reference 90]) excluded a pentamer and dictated a hexameric or tetrameric structure for the cereal germins. The M_r (~25) of the glycosylated germin monomer, based on its mobility in sodium dodecyl sulfate-polyacrylamide gel electrophoresis gels, had been incorrect owing to a glycan-induced anomaly. The correct M_r of the germin monomer, based on the sizes of its polypeptide and *N*-glycan constituents (20 + 2=22) (135), conforms with sedimentation-equilibrium measurements of the M_r (~130) only if germin is homohexameric, not homotetrameric. This conclusion was confirmed, definitively, by a more comprehensive X-ray diffraction study of barley OXO crystals in our own laboratory (310); this enzyme contains a hexameric arrangement of subunits of the type found in the storage proteins, but of course these latter proteins are composed of a trimer of two-domain subunits rather than being a trimer of single-domain dimers. The homology model of OXO (90) also confirmed the potential catalytic significance and metal-binding capacity of the three conserved His residues located within motifs 1 and 2 at the center of the β-barrel of many cupins. It is now considered likely that such a His cluster, together with an adjacent conserved Glu, may be the binding site for an Mn^{2+} ion recently found to be the metal present in OXO, at least in those isolated from cereals (39, 238, 239). A similar combination of modelling and experimental approaches could be made using the structural data from the sugar-binding domain of the bacterial AraC transcription factor (270) and the sequence data from SBP (bicupins) from higher plants (40, 219) in order to identify ligands specifically involved in the binding of either mono- or disaccharides in these subgroups.

SUMMARY OF CUPIN FUNCTIONS

In conclusion, therefore, cupins are found in a wide range of cell types and have a wide range of biochemical functions, including several enzymes related to cell wall synthesis, particularly in reactions involving sugar modification. There seems to be a consistent association with stress responses in both prokaryotes and eukaryotes. In higher plants, this association with desiccation tolerance (20) is exemplified by the seed storage proteins, a specialized and well-characterized group of nonenzymatic proteins; they contain at most a single conserved His, and no enzyme activity is known. One additional recurring theme in this analysis of cupin function is the link to oxalate metabolism, an area of biochemistry that has received little attention recently. The section below will attempt to remedy this omission by emphasizing the significance of oxalate (and oxalate degradation) in several fields of microbiology, plant science, food science, and medicine.

BIOLOGICAL SIGNIFICANCE OF CUPINS IN OXALATE METABOLISM

There are two pieces of important evidence that suggest a link between some of the cupin proteins and oxalate metabolism. First, the archetypal member of the cupin family is wheat germin, a cereal protein with OXO activity (173) (see above). Second, and most significantly, the fungal OXDCs have been found previously to be bicupin proteins (69). It is very likely, therefore, that other cupins have similar enzymatic activities, and in particular one would predict that the two 20+20 bicupins from *B. subtilis* and the related sequence from *Streptococcus mutans* identified in this review (Fig. 3) are also OXDCs. Unfortunately, there is scant information about the role of oxalate metabolism in these bacterial species. It is known, however, that oxalic acid is among the range of organic acids produced by strains of *B. subtilis* isolated from certain Indian soils (17).

Microbiological Significance of Oxalic Acid and Oxalate-Degrading Enzymes

Oxalic acid has been implicated in a wide range of environmental effects including several biological and geochemical processes in soils (71, 101). For example, oxalate is the major organic anion in many forest soils throughout the world, and as such it has a large effect on the availability of phosphorus, aluminum, and calcium (154). This is because oxalate will chelate aluminum and iron, thereby making more phosphorus available to plant roots. Interestingly, aluminum has recently been found to stimulate the production and secretion of oxalate from roots of buckwheat (188, 323) in addition to inducing the synthesis of a defence related GLP in wheat. As well as this role in plant nutrition, oxalic acid is linked to general weathering of soil minerals and the subsequent precipitation of insoluble metal oxalates (87). This latter process is associated with the survival of fungi growing in the presence of potentially toxic metal compounds (e.g., copper-containing wood preservatives). It has also been exploited in the solubilization of heavy metals from bauxite, clay, sand, sewage sludge, and other metal bearing materials. In many of these applications, *Aspergillus niger* is favored as the best species for oxalic acid production (251, 252, 279), in contrast to the medical dangers of such production by this organism.

A related environmental role for oxalic acid (and oxalate-degrading enzymes) comes from its key importance in the carbon cycle and the release of CO_2 from rotting wood (98, 99), a process largely mediated by basidiomycetous white rot fungi. Such fungi, along with their brown rot equivalents (203), are known to produce oxalic acid (72, 257) and oxalate-degrading enzymes (73, 202) under some conditions. The biochemical

or the role of oxalate include its capacity to chelate nese and thus to stimulate the activity of Mn peroxidase, acellular, glycosylated enzyme involved in lignin degra- (16, 163, 292). This enzyme is also implicated in the ercial use of the fungus *Bjerkandera* in the biobleaching gnified craft pulp during paper making (201, 208, 236). pplications of this type, the level of oxalate, and thus the al value of the particular technology, depends upon the f the OXDC (or equivalent) enzyme, and therefore the ery of the putative structure of OXDC, albeit indirectly ue of its relationship to OXO and seed storage proteins provides the first opportunity to consider directed mod- on of the enzyme.

; also relevant to emphasize the key role of manganese in emical reactions utilized in both lignin synthesis (Mn^{2+} : metal ion present in the active site of OXO, which des H_2O_2 required for cell wall cross-linking) (239) and degradation (see the comments above on lignin peroxi- . Oxalate and manganese together therefore control a segment of the carbon cycle which involves both the on of carbon in woody biomass and its later release from material.

Role of Oxalate in Plant Pathogenesis

; described below in the context of transgenic plants, oxalic is a toxin associated with many plant pathogens, particu- *Sclerotinia sclerotiorum* (185) and related species. It is ted into the plant during the infection process, is impli- d in the degradation of the plant cells, and is then often ed by the pathogen or precipitated in the form of calcium ate crystals (78, 315).

COMMERCIAL SIGNIFICANCE OF OXALATE-DEGRADING ENZYMES

everal of the oxalate-degrading enzymes of bacterial, fun- (52; Scelonge and Bidney, PCT patent application WO 12827), or plant (G. Freyssinet and A. Sailand, 17 Septem- 1992, PCT patent application WO 92/15685) origin, many them cupins, have either actual or potential commercial nificance, with applications in many areas of medicine, ag- ulture, and industry. This section reviews these uses and nonstrate how the increasing understanding of the cupin perfamily will have practical benefits.

Medical Diagnosis and Treatment

As described above, the OXO and OXDC enzymes have eat commercial importance in assays for oxalate in blood and ine (116, 123). Such assays are vital in the control of hyper- aluria, particularly in patients suffering from urolithiasis, i.e., e formation of stones (crystals of calcium oxalate) in the inary tract (245); this complaint afflicts more than 10% of the .S. population. Excess oxalate levels in humans can have ny causes, including hereditary intolerance to the normal vels of oxalate in the diet (100); consumption of ascorbate .3), ethylene glycol, diethylene glycol, or xylitol; inhalation of ie anaesthetic methoxyflurane; and infection with *Aspergillus* nown as aspergillosis) (169, 226). Such levels are also found n cystic fibrosis patients as a consequence of antibiotic treat- ent that eliminates *Oxalobacter formigenes* from the gut flora 262) and in premature babies fed with infant formula rather han breast milk (124). In addition to dietary or pharmacolog- al treatments to reduce metabolic oxalate levels (120, 121), ther suggested means of reducing oxalate intake include the se of immobilized or encapsulated (18, 19) oxalate-degrading enzymes in the digestive tract or in the peritoneal cavity. For example, OXO which had been immobilized by adsorption onto chitosan and cross-linking with glutaraldehyde was shown to have enhanced resistance to proteolytic digestion and heavy-metal inactivation and was considered suitable for oral administration (231). In a related study (230), rats implanted in their peritoneal cavity with dialysis membranes containing banana OXO were able to metabolize intraperitoneally injected [^{14}C]oxalate, as well as its precursor, [^{14}C]glyoxalate. The leading therapeutic compound in preclinical development is probably IxC2-62/47 (from Ixion Biotechnology Inc.), an orally administered product consisting of two recombinant enzymes of bacterial origin (M. J. Allison and H. Sidhu, 26 November 1998, PCT patent application WO 98/52586).

As an alternative to treatment of patients with these conditions, it is possible to reduce the oxalate content of food prior to its being eaten (176; R. Yoshida, T. Tanaka, and Y. Hotta, 5 March 1994, European patent application 0639329 A1). In one such method, it was shown (132) that the oxalate-degrading bacterium *Eubacterium lentum* WYH-1, isolated from human feces (133), could completely degrade 1 mg of oxalate per ml in artificial intestinal juice or the oxalate in an infusion of black tea—one of the major sources of oxalate in the human diet (321). Similarly, barley roots in a stirred-tank reactor have been used to reduce the level of oxalate in ginger juice (313). The use of microorganisms or isolated enzymes for this purpose is the subject of several patent applications (G. Kohlbecker, F. Heinz, and R. Schildbach, 8 April 1982, German patent application). Transgenic approaches to reducing oxalate levels in plants will be considered below.

In addition to this relatively common metabolic disorder, there is a more infrequent but much more serious genetic condition (primary hyperoxaluria), which leads to deposition of calcium oxalate throughout the body, often with fatal consequences. Treatment of this condition is also discussed below.

Human Gene Therapy

As mentioned in the previous section, primary hyperoxaluria (55) is a potentially fatal condition. It exists in two forms, type 1, which is due to a defect in the liver-specific peroxisomal enzyme alanine:glyoxylate aminotransferase/serine pyruvate aminotransferase, and type 2, which involves glyoxalate reductase/D-glycerate dehydrogenase, a cytosolic enzyme found in hepatocytes and leukocytes. At present, treatments include pharmacological approaches (121) or preemptive liver transplantation (149, 254). Aside from the implantation of immobilized OXO or the oral administration of recombinant enzymes, there has been a longstanding interest in the potential of gene therapy for treatment of these life-threatening conditions. Although their eventual aim was to clone and utilize the fungal oxalate decarboxylase gene (263; B. Finlayson and A. Peck, 3 November 1988, PCT patent application), success was first achieved in the cloning of the bacterial equivalent (186, 187), the oxalyl coenzyme A decarboxylase from *Oxalobacter formigenes*, an anaerobic oxalate-degrading bacterium found in the mammalian intestine (4, 15, 165). It is hoped that the increased understanding of the oxalate-degrading enzymes described in this review will soon lead to further advances in this important area of human therapy.

Transgenic Plants

Resistance to plant pathogens. Possibly the most extensive application of oxalate-degrading enzymes concerns their use in improving the tolerance of plants to fungal pathogens. The basis of this strategy is the toxic effect of the oxalic acid se-

creted by many such pathogens, particularly *Sclerotinia sclerotiorum* (229), a fungus with a wide host range including sunflower (195) and oilseed rape. During severe outbreaks crop losses can reach 60%. At present, control of the disease by fungicide application to the plant is expensive and not always reliable. The overwintering sclerotia can remain dormant in the soil for several years, before germinating and either producing apothecia (fruiting bodies which generate large numbers of the infective ascospores) or growing directly into mycelia, which can invade the plant stem at ground level. Few genetic sources of resistance to the pathogen are available to the plant breeder, and despite the potential of mycoparasites as a means of biological control (306) there is a continuing demand for new approaches to combat this disease (246).

It has been known for several years that the mode of action of this pathogen involves an important role for oxalic acid (215, 244, 325). This acid is secreted by the fungal mycelium, which either develops on the petal and leaf surfaces (138) after germination of the ascospore or grows out directly from the perennating sclerotium after its germination in the soil. As the pathogen-derived acid enters the plant, it chelates the calcium from the middle lamellae of the plant cell walls, thus inducing embolisms in the xylem vessels and causing wilting (272). The acid also inhibits the activity of *o*-diphenol oxidase activity in host cells (76), thereby suppressing defense responses. In addition, it reduces the internal pH of the plant and consequently stimulates the activity of the cell wall-degrading cellulases and pectinases produced by the pathogen (192). Genetic evidence for the role of oxalic acid comes from studies on mutant strains of the fungus which are deficient in oxalate production and also avirulent; revertants regain their virulence (95).

A strategy was therefore developed to introduce into susceptible plants a gene encoding an oxalate-degrading enzyme which could reduce the level of the pathogen-derived toxin and thereby reduce the growth of the infective mycelium and spread of the disease. The best results achieved to date (287, 288; D. L. Bidney, D. G. Charne, S. L. Coughlan, I. Falak, M. K. Mancl, K. A. Nazarian, C. J. Scelonge, and N. Yalpani, 28 January 1999, PCT patent application; C. Thompson, G. Nisbet, H. Jones, and J. M. Dunwell, unpublished observations), refer to investigations with oilseed rape and sunflower transformed with the barley or wheat OXO. Such results have considerable commercial potential (82, 111). Additionally, it has been proposed that oxalate-degrading genes could be used as selectable markers in transformation experiments (82, 253).

Improvements in digestibility. Oxalic acid and its salts are well known to be toxic to humans at high doses (104, 183) and can cause medical problems even when present at low concentrations in the diet. Certain leafy vegetables such as spinach and *Amaranthus* have particularly high concentrations, and there are many breeding efforts to improve the palatability of these and other crops for humans, and also for farm ruminants such as sheep and goats where aversion to fodder has been associated with oxalate content (83, 167).

In addition to a general reduction in levels of soluble or insoluble oxalate, it is possible that introduction of oxalate-degrading genes (199) may be useful as a means of reducing specific oxalate-related toxins such as the well-known neurotoxin β-*N*-oxalyl-1-α,β-diaminopropanoic acid (ODAP), a compound that is found in the legume species *Lathyrus sativus* and is associated with the development of lathyrism (233, 283), a severe neurological disease.

Bioremediation and Industrial Uses

Oxalate, particularly sodium oxalate, is a significant environmental contaminant in industrial processes such as the smelting of alumina by the Bayer process. At present, the disposal methods comprise either burning or burial in landfill sites. It is hoped that these methods may be replaced in future by use of various means of bacterial degradation involving either an alkalophilic *Bacillus* species (209) or *Pseudomonas oxalaticus* (11), an organism isolated from rhubarb patches (a rich source of oxalate). It is possible that the recently described new species of obligately oxalotrophic *Ammonophilus oxalaticus* and *A. oxalivorans* (319) isolated from the rhizosphere of sorrel (*Rumex acetosa*, another species with a high oxalate content) could be used for these purposes.

A recent application in this area is the claimed use (N. O. Nilvebrandt, A. Reiman, and F. De Sousa, 26 February 1998, PCT patent application) of oxalate-degrading enzymes (OXO and/or OXDC) to reduce the levels of oxalate in the process liquids during the production of pulp and paper. Wood contains oxalic acid at concentrations of 0.1 to 0.4 kg/ton (bark contains up to 15 kg/ton), and during processing this compound can easily precipitate in the form of calcium oxalate crystals, which cause problems in pipework, washing filters, and heat exchangers. A similar previous application (117) involved the use of these enzymes to inhibit the deposition of beer stones during the brewing process. Levels of oxalate also must be kept low to prevent the problem of "gushing" (see also reference 107) during this process.

Another potential industrial use of OXDC and related enzymes concerns the use of fungal exopolysaccharides. One such compound is scleroglucan, a neutral glucan produced by *Sclerotinia glucanicum* and composed of a linear chain of β-D(1,3)-linked D-glucopyranosyl residues with single D-glycopyranosyl residues linked β(1,6) to every third residue of the main chain. Because of its structure and high molecular weight, it has great value as a viscosifier in enhanced oil recovery, a role it which its closest rival is xanthan gum (122). Unfortunately, production of scleroglucan in reactors is accompanied by the concurrent synthesis of oxalate (302), an unwanted by-product that makes purification more costly. Such synthesis is stimulated at a pH above 3.5, partly because of inhibition of OXDC. It would therefore be most valuable to be able to modify the characteristics of this enzyme in order to reduce the level of oxalate formed during the production process.

OXALATE AND THE ORIGIN OF LIFE

In a recent review (112) of photosynthesis and the origin of life, it was suggested that the first stage in the development of photosynthesis took place in iron-rich clays and involved the photoreduction of carbon dioxide to form oxalate, which was then reduced to glyoxalate with the aid of manganese (see the comments on Mn^{2+} as the active-site metal in barley OXO above). This phase was then followed by the entry of sulfur into the evolving clay systems, the subsequent ability to fix nitrogen, and, finally, the involvement of phosphate. Oxalate is therefore implicated in the primary production of all organic chemicals on Earth.

ORIGINAL FUNCTION OF THE ANCESTRAL "PROTOCUPIN"

It is possible that the ancestral "protocupin" was a small protein (ca. 100 residues with an intermotif spacing of 15 aa) that was very heat stable (cf. proteins gi|2984227 from *Aquifex*

olicus and gi|2128971 from *Methanococcus jannaschii*, with optimum growth temperatures of 90 and 95°C, respectively), was probably metal containing, and had a relatively nonspecific ability to bind a range of molecules which included sugar residues. This progenitor protein then assumed other functions, for example as a PMI able to bind mannose-6-phosphate, and was also involved in a fusion event whereby a sugar-binding form of the cupin domain combined with a DNA-binding domain to generate the multidomain AraC type of transcription factor, in which the sugar-binding element is located within the N-terminal effector-binding domain (270). (In this context it is relevant to note the close association of wheat OXO with the arabinose-rich hemicelluloses [135].) Further diversification of function of the original cupin domain then followed during subsequent evolution, by insertion of residues in the intermotif region (Fig. 1) (168), addition of residues at both ends of the protein (probably associated with the generation of a multimeric structure), and duplication or fusion of the entire sequence to produce bicupins. In addition, many of the eukaryotic cupin genes have introns, and this may be associated with the occurrence of exon shuffling during the evolutionary process. Such variation means that the underlying cupin structure is now found in proteins that range in size from the extremophile proteins of about 100 aa to the heat-stable, hexameric, manganese-containing OXO from cereals (201 aa in each subunit [310]) and the desiccation-tolerant, trimeric seed storage proteins (400 to 500 aa in each two-domain unit) from higher plants. It is hoped that an improved understanding of the structure-function relationships in this diverse superfamily of proteins will soon help to elucidate the factors that allowed such a diversity to develop during the evolutionary process and led eventually to the proteins that are now the major part of the human diet.

In the context of this evolutionary summary, it should be noted that the sizes of the two conserved motifs within the cupin signature (20 or 21 aa in motif 1, and 16 aa in motif 2) correspond quite closely to the sizes of two of the conserved units (15, 22, and 30 aa) presumed to make up all ancestral proteins in the progenote (56, 57, 94) and to be encoded by the ancestral exons.

CONCLUDING REMARKS AND FUTURE DIRECTIONS

The discovery of the cupin superfamily epitomizes the value of an integrated approach to the study of protein structure and function (24, 38, 119, 286). Only by such integration of information available from the previous X-ray analysis of storage proteins (155, 177) with data from the prokaryotic genome programs was it possible to detect the evolutionary connection between this wide spectrum of enzymes and binding proteins and thereby to reinforce the concept that there is only a relatively small number of structurally conserved elements from which all proteins are constructed (271). This new form of biological research, which perhaps could be designated protein molecular biology, is a direct consequence of the increasing power of analytical algorithms designed to compare primary sequence information with 3D structural data. In this particular example of multifunctional cupins (64, 69), a discovery recently validated by Aravind and Koonin (12), we have uncovered an evolutionary sequence in which a particularly stable β-sheet structure has been coopted for a variety of purposes, many or all of which require a thermostable (222, 223), pepsin-resistant (170) framework, usually containing metal-binding ligands (90, 239). It is very likely that similar conserved ligands in different members of the cupin family will be found to bind different metals, a situation reminiscent of that found in the homoprotocatachuate 2,3-dioxygenases.

Future studies will take a number of directions. First, they will certainly add to this preliminary survey of superfamily members as more genomes are completed. More importantly, the functional information obtained from prokaryotic studies will allow rational predictions of function for the variety of related proteins in eukaryotes. Undoubtedly, this process will be aided by the increasingly efficient microarray/chip technology (61) that is able to illustrate the pattern of expression of all genes within a single organism grown in one or more environments. In particular, it is hoped that eventually the role of the large number of GLPs in higher plants can be exemplified. As described above, many of these proteins are apoplastic and are associated with responses to biotic and abiotic stress. They are therefore of prime interest to plant breeders interested in improving the growth of crops under limiting environments. In a more fundamental context, resolution of additional cupin structures at the atomic level (177, 270, 310) (Fig. 2) will confirm the exact significance of the key active-site residues and those that confer molecular integrity under extremes of temperature and in the presence of normally destructive chemicals. Such information will prove of great potential value, for example in applied agricultural, medical, or environmental projects utilizing oxalate-degrading enzymes and in nutritional projects that aim to improve the quality of seed proteins.

ACKNOWLEDGMENTS

J.M.D. thanks the Biotechnology and Biological Sciences Research Council (BBSRC) and Zeneca Ltd. for financial support during the course of this review. S.K. is the recipient of a BBSRC Resource and Stress Allocation in Plants grant.

Thanks are due to the two reviewers, who provided most detailed and helpful recommendations.

REFERENCES

1. Adams, P., J. C. Thomas, D. M. Vernon, H. J. Bohnert, and R. G. Jensen. 1992. Distinct cellular and organismic responses to salt stress. Plant Cell Physiol. 33:1215–1223.
2. Allen, R. G., R. K. Newton, K. J. Farmer, and C. Nations. 1985. Effects of the free radical generator paraquat on differentiation, superoxide dismutase, glutathione and inorganic peroxides in microplasmodia of *Physarum polycephalum*. Cell Tissue Kinet. 18:623–630.
3. Allen, R. G., R. K. Newton, R. S. Sohal, G. L. Shipley, and C. Nations. 1985. Alterations in superoxide dismutase, glutathione, and peroxides in the plasmodial slime mold *Physarum polycephalum* during differentiation. J. Cell. Physiol. 125:413–419.
4. Allison, M. J., S. L. Daniel, and N. A. Cornick. 1995. Oxalate degrading bacteria, p. 131–168. *In* S. R. Khan (ed.), Calcium oxalate in biological systems. CRC Press, Inc., Boca Raton, Fla.
5. Reference deleted.
6. Allona, I., M. Quinn, E. Shoop, K. Swope, S. S. Cyr, J. Carlis, J. Riedl, E. Retzel, M. M. Campbell, R. Sederoff, and R. W. Whetten. 1998. Analysis of xylem formation in pine by cDNA sequencing. Proc. Natl. Acad. Sci. USA 95:9693–9698.
7. Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programmes. Nucleic Acids Res. 25:3389–3402.
8. Alvarez, M. A., H. Fu, C. Khosla, D. A. Hopwood, and J. E. Bailey. 1996. Engineered biosynthesis of novel polyketides: properties of the *whiE* aromatase/cyclase. Nat. Biotechnol. 14:335–338.
9. Andersson, S. G. E., A. Zomorodipour, J. O. Andersson, T. Sicheritz-Pontén, U. C. M. Alsmark, R. M. Podowski, A. K. Näslund, A-S. Eriksson, H. H. Winkler, and C. G. Kurland. 1998. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. Nature 396:133–140.
10. Andolfatto, P., A. Bornhauser, H. J. Bohnert, and R. G. Jensen. 1994. Transformed hairy roots of *Mesembryanthemum crystallinum*: gene expression patterns upon salt stress. Physiol. Plantarum 90:708–714.
11. Anonymous. 1992. These rhubarb-feeding bugs like sodium oxalate too. Chem. Eng. Oct:17.
12. Aravind, L., and E. V. Koonin. 1999. Gleaning non-trivial structural, functional and evolutionary information about proteins by iterative database

searches. J. Mol. Biol. 287:1023–1040.
13. Auer, B. L., D. Auer, and A. L. Rodgers. 1998. Relative hyperoxaluria, crystalluria and haematuria after megadose ingestion of vitamin C. Eur. J. Clin. Investig. 28:695–700.
14. Aurora, R., and G. D. Rose. 1998. Seeking an ancient enzyme in *Methanococcus jannaschii* using ORF1, a protein based on predicted secondary structure comparisons. Proc. Natl. Acad. Sci. USA 95:2818–2823.
15. Baetz, A. L., and M. J. Allison. 1989. Purification and characterisation of oxalyl-coenzyme A decarboxylase from *Oxalobacter formigenes*. J. Bacteriol. 171:2605–2608.
16. Bunci, L., I. Bertini, L. Dal Pozzo, R. Del Conte, and M. Tien. 1998. Monitoring the role of oxalate in manganese peroxidase. Biochemistry 37:9009–9015.
17. Banik, S., and B. K. Dey. 1983. Phosphate-solubilizing potentiality of the microorganisms capable of utilizing aluminium phosphate as a sole phosphate source. Zentbl. Mikrobiol. 138:17–23.
18. Batich, C., and F. Vaghefi. February 1994. Process for microencapsulating cells. U.S. patent 5,286,495.
19. Batich, C., and F. Vaghefi. July 1997. Process for microencapsulating cells. U.S. patent 5,648,099.
20. Bäumlein, H., H. Braun, I. A. Kakhovskaya, and A. D. Shutov. 1995. Seed storage proteins of spermatophytes share a common ancestor with desiccation proteins of fungi. J. Mol. Evol. 41:1070–1075.
21. Beagles, K. E., P. F. Morrison, and M. P. Heyes. 1998. Quinolinic acid *in vivo* synthesis rates, extracellular concentrations, and intercompartmental distributions in normal and immune-activated brain as determined by multi-isotope microdialysis. J. Neurochem. 70:281–291.
22. Becker, A., F. Katzen, A. Puhler, and L. Ielpi. 1998. Xanthan gum biosynthesis and application: a biochemical/genetic perspective. Appl. Microbiol. Biotechnol. 50:145–152.
23. Becker, W. M. 1997. Characterization of Ara h1 by two-dimensional electrophoresis immunoblot and recombinant techniques: new digestion experiments with peanuts imitating the gastrointestinal tract. Int. Arch. Allergy Immunol. 113:118–121.
24. Bergdoll, M., L. D. Eltis, A. D. Cameron, P. Dumas, and J. T. Bolin. 1998. All in the family: structural and evolutionary relationships among three modular proteins with diverse functions and variable assembly. Protein Sci. 7:1661–1670.
25. Bergh, S., and M. Uhlén. 1992. Analysis of a polyketide synthesis-encoding gene cluster of *Streptomyces curacoi*. Gene 117:131–136.
26. Bergh, S., and M. Uhlén. 1992. Cloning, analysis, and heterologous expression of a polyketide synthase gene cluster of *Streptomyces curacoi*. Biotechnol. Appl. Biochem. 15:80–89.
27. Berna, A., and F. Bernier. 1997. Regulated expression of a wheat germin gene in tobacco: oxalate oxidase activity and apoplastic localization of the heterologous protein. Plant Mol. Biol. 33:417–429.
28. Berna, A., and F. Bernier. 1999. Regulation by biotic and abiotic stress of a wheat germin gene encoding oxalate oxidase, a H_2O_2-producing enzyme. Plant Mol. Biol. 39:539–549.
29. Bernier, F., G. Lemieux, and A. Paiotta. 1987. Gene families encode the major encystment proteins of *Physarum polycephalum* plasmodia. Gene 59:265–277.
30. Bernier, F., A. Paiotta, and G. Lemieux. 1986. Molecular cloning of mRNAs expressed specifically during sphérulation in *Physarum polycephalum*. Biochim. Biophys. Acta 867:234–243.
31. Bernier, F., V. L. Selgy, D. Paiotta, and G. Lemieux. 1986. Changes in gene expression during sphérulation in *Physarum polycephalum*. Biochem. Cell Biol. 64:337–343.
32. Bhandari, A., D. G. Jones, J. R. Schullek, K. Vo, C. A. Schunk, L. L. Tamannha, D. Chen, Z. Yuan, M. C. Needels, and M. A. Gallop. 1998. Exploring structure-activity relationships around the phosphomannose isomerase inhibitor AF14049 via combinatorial synthesis. Bioorgan. Med. Chem. Lett. 8:2303–2308.
33. Bibb, M. J., S. Biro, H. Motamedi, J. F. Collins, and C. R. Hutchinson. 1989. Analysis of the nucleotide sequence of the *Streptomyces glaucescens* *tcm1* genes provides key information about the enzymology of polyketide antibiotic synthesis. EMBO J. 8:2727–2736.
34. Reference deleted.
35. Bjork, S. K. M., B. K. Gotthammar, M. T. Linderberg, J. P. Luthman, K. M. I. Persson, and R. Schwarcz. August 1997. 3-hydroxy anthanilic acid derivatives. U.S. patent 5,661,183.
36. Blanco, G., P. Brian, A. Peredа, C. Méndez, J. A. Salas, and K. F. Chater. 1993. Hybridization and DNA sequence analyses suggest an early evolutionary divergence of related biosynthetic gene sets encoding polyketide antibiotics and spore pigments in *Streptomyces* spp. Gene 130:107–116.
37. Bohnert, H. J., J. A. Ostrem, J. C. Cushman, C. B. Michalowski, J. Rickers, G. Meyer, E. J. Derocher, D. M. Vernon, M. Krueger, L. Vasquez-Martin, J. Velten, R. Hoefner, and J. M. Schmitt. 1988. *Mesembryanthemum crystallinum*, a higher plant model for the study of environmentally induced changes in gene expression. Plant Mol. Biol. Rep. 6:10–28.
38. Bork, P., and E. V. Koonin. 1998. Predicting function from protein sequences—where are the bottlenecks? Nature Genet. 18:313–318.
39. Reference deleted.
40. Braun, H., A. Catbal, A. D. Shutov, and H. Bäumlein. 1996. A vicilin-like seed protein of cycads: similarity to sucrose-binding proteins. Plant Mol. Biol. 31:35–44.
41. Brown, J. C., and A. M. Jones. 1994. Mapping the auxin-binding site of auxin-binding protein 1. J. Biol. Chem. 269:21136–21140.
42. Bult, C. J., O. White, G. J. Olsen, L. Zhou, R. D. Fleischmann, G. G. Sutton, J. A. Blake, L. M. Fitzgerald, R. A. Clayton, J. D. Gocayne, A. R. Kerlavage, B. A. Dougherty, J.-F. Tomb, M. D. Adams, C. I. Reich, R. Overbeek, E. F. Kirkness, K. G. Weinstock, J. M. Merrick, A. Glodek, J. L. Scott, N. S. M. Geoghagen, J. F. Weidman, J. L. Fuhrman, E. A. Presley, D. Nguyen, T. R. Utterbeck, J. M. Kelley, J. D. Peterson, P. W. Sadow, M. C. Hanna, M. D. Cotton, M. A. Hurst, K. M. Roberts, B. P. Kaine, M. Borodovsky, H.-P. Klenk, C. M. Fraser, H. O. Smith, C. R. Woese, and J. C. Venter. 1996. Complete genome sequence of the methanogenic archaeon *Methanococcus jannaschii*. Science 273:1058–1073.
43. Burks, A. W., D. Shin, G. Cockrell, J. S. Stanley, R. M. Helm, and G. A. Bannon. 1997. Mapping and mutational analysis of the IgE-binding epitopes on Ara h 1, a legume vicilin protein and a major allergen in peanut hypersensitivity. Eur. J. Biochem. 15:334–339.
44. Bustos, S. A., and R. F. Schleif. 1993. Functional domains of the AraC protein. Proc. Natl. Acad. Sci. USA 90:5638–5642.
45. Caliskan, M., and A. C. Cuming. 1998. Spatial specificity of H_2O_2-generating oxalate oxidase gene expression during wheat embryo germination. Plant J. 15:165–171.
46. Carter, C., R. A. Graham, and R. W. Thornburg. 1998. *Arabidopsis thaliana* contains a large family of germin-like proteins: characterization of cDNA and genomic sequences encoding 12 unique family members. Plant Mol. Biol. 38:929–943.
46a. Carter, C., R. A. Graham, and R. W. Thornburg. 1999. Nectarin I is a novel soluble germin-like protein expressed in the nectar of *Nicotiana* sp. Plant Mol. Biol. 41:207–216.
47. Chet, I., and H. P. Rusch. 1969. Induction of spherule formation in *Physarum polycephalum* by polyols. J. Bacteriol. 100:674–676.
48. Cole, S. T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. Barry, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, S. Oliver, J. Osborne, M. A. Quail, M. A. Rajandream, J. Rogers, S. Rutter, K. Seeger, S. Skelton, R. Squares, S. Squares, J. E. Sulston, K. Taylor, S. Whitehead, and B. G. Barrell. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature 393:537–544.
49. Corpet, F., J. Gouzy, and D. Kahn. 1998. The ProDom database of protein domain families. Nucleic Acids Res. 26:323–326.
50. Crawford, R. L., S. W. Hutton, and P. J. Chapman. 1975. Purification and properties of gentisate 1,2-dioxygenase from *Moraxella osloensis*. J. Bacteriol. 121:794–799.
51. Davis, N. K., and K. F. Chater. 1990. Spore colour in *Streptomyces coelicolor* A3(2) involves the developmentally regulated synthesis of a compound biosynthetically related to polyketide antibiotics. Mol. Microbiol. 4:1679–1691.
52. Datta, A., A. Mehta, and K. Natarajan. August 1996. Oxalate decarboxylate. U.S. patent 5,547,870.
53. Deckert, G., P. V. Warren, T. Gaasterland, W. G. Young, A. L. Lenox, D. E. Graham, R. Overbeek, M. A. Snead, M. Keller, M. Aujay, R. Huber, R. A. Feldman, J. M. Short, G. J. Olsen, and R. V. Swanson. 1998. The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. Nature 392:353–358.
54. De Jong, E. C., M. Van Zijverden, S. Spanhaak, S. J. Koppelman, H. Pellegrom, and A. H. Penninks. 1998. Identification and partial characterization of multiple major allergens in peanut proteins. Clin. Exp. Allergy 28:743–751.
55. De Pauw, L., and C. Toussinet. 1996. Primary hyperoxaluria. Rev. Med. Brux. 17:67–74.
56. De Souza, S. J., M. Long, L. Schoenbach, S. W. Roy, and W. Gilbert. 1997. The correlation between introns and the three-dimensional structure of proteins. Gene 31:141–144.
57. De Souza, S. J., M. Long, R. J. Klein, S. Roy, S. Lin, and W. Gilbert. 1998. Towards a resolution of the introns early/late debate: only phase zero introns are correlated with the structure of ancient proteins. Proc. Natl. Acad. Sci. USA 95:5094–5099.
58. Dixon, D. C., J. R. Catt, and D. F. Klessig. 1991. Differential targeting of the tobacco PR-1 pathogenesis-related proteins to the extracellular space and vacuoles of crystal idioblasts. EMBO J. 10:1317–1324.
59. Domon, J.-M., B. Dumas, E. Lainé, Y. Meyer, A. David, and H. David. 1995. Three glycosylated polypeptides secreted by several embryogenic cell lines of pine show highly specific serological affinity to antibodies directed against the wheat germin apoprotein monomer. Plant Physiol. 108:141–148.
60. Doten, R. C., and R. P. Mortlock. 1985. Inducible xylitol dehydrogenases in enteric bacteria. J. Bacteriol. 162:845–848.

D. J., M. Bittner, Y. Chen, P. Meltzer, and J. M. Trent. 1999. ion profiling using cDNA microarrays. Nat. Genet. 21:10–14.
B., G. Freyssinet, and K. E. Pallett. 1995. Tissue specific expression in-like oxalate oxidase during development and fungal infection of seedlings. Plant Physiol. 107:1091–1096.
. B., A. Sailland, J-P. Cheviet, G. Freyssinet, and K. Pallett. 1993. cation of barley oxalate oxidase as a germin-like protein. C. R. Acad. ris 316:793–798.
ll, J. M. 1998. Cupins: a new superfamily of functionally diverse ns that include germins and plant storage proteins. Biotechnol. . Eng. Rev. 15:1–32.
ll, J. M. 1998. Sequence analysis of the cupin gene family in *Syn-atis* PCC6803. Microb. Comp. Genom. 3:141–148.
ell, J. M. 1998. How to engineer a crop plant. Pestic. Outlook 9:29–33.
ell, J. M. 1998. Novel food products from genetically modified crop s: methods and future prospects. Int. J. Food Sci. Technol. 33:205–213.
ence deleted.
ell, J. M., and P. J. Gane. 1998. Microbial relatives of seed storage ins: conservation of motifs in a functionally diverse superfamily of mes. J. Mol. Evol. 46:147–154.
ag, F., J. Eichler, A. Price, M. R. Leonard, and W. Wickner. 1997. genesis of the gram-negative bacterial envelope. Cell 91:567–573.
ton, M. V., and C. S. Evans. 1996. Oxalate production by fungi: its role athogenicity and ecology in the soil environment. Can. J. Microbiol. 81–895.
ton, M. V., C. S. Evans, P. T. Atkey, and D. A. Wood. 1993. Oxalate duction by Basidiomycetes, including the white-rot species *Coriolus ver-lor* and *Phanerochaete chrysosporium*. Appl. Microbiol. Biotechnol. 39: 0.
tton, M. V., M. Kathiara, L. M. Gallagher, and C. S. Evans. 1994. rification and characterization of oxalate decarboxylase from *Coriolus ricolor*. FEMS Microbiol. Lett. 116:321–326.
lls, L. D., and J. T. Bolin. 1996. Evolutionary relationships among extra-ol dioxygenases. J. Bacteriol. 178:5930–5937.
spelund, M., S. Saeboe-Larsson, D. W. Hughes, G. A. Galau, F. Larsen, d K. S. Jakobsen. 1992. Late embryogenesis-abundant genes encoding roteins with different numbers of hydrophilic repeats are regulated differ-ntially by abscisic acid and osmotic stress. Plant J. 2:241–252.
errar, P. H., and J. R. L. Walker. 1993. O-Diphenol oxidase inhibition—an dditional role for oxalic acid in the phytopathogenic arsenal of *Sclerotinia clerotiorum* and *Sclerotium rolfsii*. Physiol. Mol. Plant Pathol. 43:415–427.
Reference deleted.
Francesehi, V. R., and F. A. Loewus. 1995. Oxalate biosynthesis and func-tion in plants and fungi. p. 113–130. *In* S. R. Khan (ed.), Calcium oxalate in biological systems. CRC Press, Boca Raton, Fla.
Fraser, C. M., S. Casjens, W. M. Huang, G. G. Sutton, R. Clayton, R. Lathigra, O. White, K. A. Ketchum, R. Dodson, E. K. Hickey, M. Gwinn, B. Dougherty, J-F. Tomb, R. D. Fleischmann, D. Richardson, J. Peterson, A. R. Kerlavage, J. Quackenbush, S. Salzburg, M. Hanson, R. van Vugt, N. Palmer, M. D. Adams, G. Gocayne, J. Weidman, T. Utterback, L. Watthey, L. McDonald, P. Artinch, C. Bowman, S. Garland, C. Fujii, M. D. Cotton, K. Horst, K. Roberts, B. Hatch, H. O. Smith, and J. C. Venter. 1997. Genomic sequence of a Lyme disease spirochaete. *Borrelia burgdorferi*. Nature 390:580–586.
Fraser, C. M., S. J. Norris, G. M. Weinstock, O. White, G. G. Sutton, R. Dodson, M. Gwinn, E. K. Hickey, R. Clayton, K. A. Ketchum, E. Sodergren, J. M. Hardham, M. P. McLeod, S. Salzberg, J. Peterson, H. Khalak, D. Richardson, J. K. Howell, M. Chidambaram, T. Utterback, L. McDonald, P. Artinch, C. Bowman, M. D. Cotton, and J. C. Venter. 1998. Complete genome sequence of *Treponema pallidum*, the syphilis spirochete. Science 281:375–388.
Freiberg, C., R. Fellay, A. Bairoch, W. J. Broughton, A. Rosenthal, and X. Perret. 1997. Molecular basis of symbiosis between *Rhizobium* and legumes. Nature 387:394–401.
Reference deleted.
Frutos, P., A. J. Duncan, I. Kyriazakis, and I. J. Gordon. 1998. Learned aversion towards oxalic acid-containing foods by goats: does rumen adaptation to oxalic acid influence diet choice? J. Chem. Ecol. 24:383–397.
Fry, S. C. 1998. Oxidative scission of plant cell wall polysaccharides by ascorbate-induced hydroxyl radicals. Eur. J. Biochem. 332:507–515.
Fu, W., and P. Oriel. 1998. Gentisate 1,2-dioxygenase from *Haloferax* sp. D1227. Extremophiles 2:439–446.
Fuenmayor, S. L., M. Wild, A. L. Boyes, and P. A. Williams. 1998. A gene cluster encoding steps in conversion of naphthalene to gentisate in *Pseudomonas* sp. strain U2. J. Bacteriol. 180:2522–2530.
Gadd, G. M. 1999. Fungal production of citric and oxalic acid: importance in metal speciation, physiology and biogeochemical processes. Adv. Microb. Physiol. 41:47–92.
Gallegos, M-T., R. Schleif, A. Bairoch, K. Hofman, and J. L. Ramos. 1997. AraC/XylS family of transcription regulators. Microbiol. Mol. Biol. Rev. 61:393–410.
Galperin, M. Y., and S. E. Brenner. 1998. Using metabolic pathway databases for functional annotation. Trends Genet. 14:332–333.
Gane, P. J., J. M. Dunwell, and J. Warwicker. 1998. Modelling based on the structure of vicilins predicts a histidine cluster in the active site of oxalate oxidase. Mol. Evol. 46:488–493.
Gaubier, P., M. Raynal, C. Hull, G. M. Huestis, F. Grellet, C. Arenas, M. Pages, and M. Delseny. 1993. Two different Em-like genes are expressed in *Arabidopsis thaliana* seeds during maturation. Mol. Gen. Genet. 238:409–418.
Geelen, D., K. Goethals, M. Van Montagu, and M. Holsters. 1995. The *nodD* locus from *Azorhizobium caulinodans* is flanked by two repetitive elements. Gene 164:107–111.
Gehring, C. A., R. M. McConchie, M. A. Venis, and R. W. Parish. 1998. Auxin-binding-protein antibodies and peptides influence stomatal opening and alter cytoplasmic pH. Planta 205:581–586.
Gilbert, W., S. J. De Souza, and M. Long. 1997. Origin of genes. Proc. Natl. Acad. Sci. USA 94:7698–7703.
Godoy, G., J. R. Steadman, M. B. Dickman, and R. Dam. 1990. Use of mutants to demonstrate the role of oxalic acid in pathogenicity of *Sclerotinia sclerotiorum* on *Phaseolus vulgaris*. Physiol. Mol. Plant Pathol. 37:179–191.
Gomes, V. M., M-I. Mosqueda, A. Blanco-Labra, M. P. Sales, K. V. S. Fernandes, R. A. Cordeiro, and J. Xavier-Filho. 1997. Vicilin storage proteins from *Vigna unguiculata* (legume) seeds inhibit fungal growth. J. Agric. Food Chem. 45:4110–4115.
Gomes, V. M., L. A. Okorokov, T. L. Rose, K. V. S. Fernandez, and J. Xavier-Filho. 1998. Legume vicilins (7S storage globulins) inhibit yeast growth and glucose stimulated acidification of the medium by yeast cells. Biochim. Biophys. Acta 1379:207–216.
Goodell, B., and J. Jellison. 1998. The role of biological metal chelators in wood degradation and in xenobiotic degradation, p. 235–249. *In* A. Bruce, and J. W. Palfreyman (ed.), Forest products biotechnology. Taylor and Francis, London, United Kingdom.
Goodell, B., J. Jellison, J. Liu, G. Daniel, A. Paszczynski, F. Fekete, S. Krishnamurthy, L. Jun, and C. Xu. 1997. Low molecular weight chelators and phenolic compounds isolated from wood decay fungi and their role in the fungal biodegradation of wood. J. Biotechnol. 53:133–162.
Goodman, H. O., R. Brommage, D. G. Assimos, and R. P. Holmes. 1997. Genes in idiopathic calcium oxalate stone disease. World J. Urol. 15:186–194.
Graustein, W. C., K. Cromack, and P. Sollins. 1997. Calcium oxalate: occurrence in soils and effect on nutrient and geochemical cycles. Science 198:1252–1254.
Griffin, A. M., E. S. Pochwijk, V. J. Morris, and M. J. Gasson. 1997. Cloning of the *accF* gene encoding the phosphomannose isomerase and GDP-pyrophosphorylase activities involved in acetan biosynthesis in *Acetobacter xylinum*. FEBS Microbiol. Lett. 154:389–396.
Grimes, H. D., P. J. Overvoorde, K. Ripp, V. R. Franceschi, and W. D. Hitz. 1992. A 62-kD sucrose binding protein is expressed and localized in tissues actively engaged in sucrose transport. Plant Cell 4:1561–1574.
Gull, J. L., I. Rodriguez-Garcia, and E. Torija. 1997. Nutritional and toxic factors in selected wild edible plants. Plant Foods Hum. Nutr. 51:99–107.
Gupta, R. S. 1997. Protein phylogenies and signature sequences: evolutionary relationships within prokaryotes and between prokaryotes and eukaryotes. Antonie Leeuwenhoek 72:49–61.
Gupta, R. S. 1998. What are archaebacteria: life's third domain or monoderm prokaryotes related to Gram-positive bacteria? A new proposal for the classification of prokaryotic organisms. Mol. Microbiol. 29:695–707.
Haas, G. J., and A. L. Fleischmann. 1961. The rapid enzymatic determination of oxalate in wort and beer. J. Agric. Food Chem. 9:451–452.
Hamel, F., C. Breton, and M. Boudet. 1998. Isolation and characterization of wheat aluminum-regulated genes: possible involvement of aluminum as a pathogenesis response elicitor. Planta 205:531–538.
Hammann, R., and H. J. Kutzner. 1998. Key enzymes for the degradation of benzoate, m- and p-hydroxybenzoate by some members of the order Actinomycetales. J. Basic Microbiol. 38:207–220.
Harpel, M. R., and J. D. Lipscombe. 1990. Gentisate 1,2-dioxygenase from *Pseudomonas acidovorans*. Methods Enzymol. 188:101–107.
Hartman, C. L., S. S. Johal, and M. R. Schmitt. February 1999. Newly characterized oxalate and uses thereof. U.S. patent 5,866,778.
Hartmann, H. 1998. Photosynthesis and the origin of life. Origins Life Evol. Biosphere 28:515–521.
Heintzen, C., R. Fischer, S. Melzer, S. Kappeler, K. Apel, and D. Staiger. 1994. Circadian oscillations of a transcript encoding a germin-like protein that is associated with cell walls in young leaves of the long-day plant *Sinapis alba* L. Plant Physiol. 106:905–915.
Henikoff, S., and J. G. Henikoff. 1991. Automated assembly of protein blocks for database searching. Nucleic Acids Res. 19:6565–6572.
Hershberger, C. L. 1996. Metabolic engineering of polyketide biosynthesis. Curr. Opin. Biotechnol. 7:560–562.
Hesse, A., D. Bangartz, H. Heynck, and W. Berg. 1996. Measurement of urinary oxalic acid: a comparison of five methods. Clin. Biochem. 29:467–472.

117. Hiatt, W. R., and J. L. Owades. March 1987. Oxalic acid removal in beer production. U.S. patent 4,652,452.
118. Hisada, S., T. Akihama, T. Endo, T. Moriguchi, and M. Omura. 1997. Expressed sequence tags of *Citrus* fruit during rapid cell development phase. 2. Cell wall. Biosci. Biotechnol. Biochem. 61:808–812.
119. Holm, L. 1998. Unification of protein families. Curr. Opin. Struct. Biol. 8:372–379.
120. Holmes, R. P., D. G. Assimos, C. D. Leaf, and J. J. Whalen. 1997. The effects of (L)-2-oxothiazolidine-4-carboxylate on urinary oxalate excretion. J. Urol. 158:34–37.
121. Holmes, R. P., and D. G. Assimos. 1998. Glyoxylate synthesis, and its modulation and influence on oxalate synthesis. J. Urol. 160:1617–1624.
122. Holzworth, G. 1985. Xanthan and scleroglucan: structure and use in enhanced oil recovery. Dev. Ind. Microbiol. 26:271–280.
123. Honow, R., D. Bongartz, and A. Hesse. 1997. An improved HPLC-enzyme-reactor method for the determination of oxalic acid in complex matrices. Clin. Chim. Acta 261:131–139.
124. Hoppe, B., B. Rath, C. Bauerfeld, and C. B. Langman. 1998. Oxalate, citrate, and sulfate concentrations in human milk compared with formula preparations: influence on urinary anion excretion. J. Pediatr. Gastroenterol. Nutr. 27:383–386.
125. Hopwood, D. A., and D. H. Sherman. 1990. Molecular genetics of polyketides and its comparison to fatty acid biosynthesis. Annu. Rev. Genet. 24:37–66.
126. Hurkman, W. J., B. G. Lane, and C. K. Tanaka. 1994. Nucleotide sequence of a transcript encoding a germin-like protein that is present in salt-stressed barley (*Hordeum vulgare* L.) roots. Plant Physiol. 104:803–804.
127. Hurkman, W. J., H. P. Tao, and C. K. Tanaka. 1991. Germin-like polypeptides increase in barley roots during salt stress. Plant Physiol. 110:971–977.
128. Hurkman, W. J., and C. K. Tanaka. 1996. Effect of salt stress on germin gene expression in barley roots. Plant Physiol. 110:971–977.
129. Hurkman, W. J., and C. K. Tanaka. 1996. Germin gene expression is induced in wheat leaves by powdery mildew infection. Plant Physiol. 111: 735–739.
130. Huynen, M. A., and P. Bork. 1998. Measuring genome evolution. Proc. Natl. Acad. Sci. USA 95:5849–5856.
131. Harslan, H., R. G. Palmer, J. Imsande, and H. T. Horner. 1997. Quantitative determination of calcium oxalate and oxalate in developing seeds of soybean (Leguminosae). Am. J. Bot. 84:1042–1046.
132. Ito, H., T. Kotake, and M. Masai. 1996. *In vitro* degradation of oxalic acid by human faeces. Int. J. Urol. 3:207–211.
133. Ito, H., N. Miura, M. Masui, K. Yamamoto, and T. Hara. 1996. Reduction of oxalate content of foods by the oxalate degrading bacterium, *Eubacterium lentum* WYH-1. Int. J. Urol. 3:31–34.
134. Iwabuchi, T., and S. Harayama. 1998. Biochemical and molecular characterization of 1-hydroxy-2-naphthoate dioxygenase from *Nocardioides* sp. KP7. J. Biol. Chem. 273:8332–8336.
135. Jaikaran, A. S. I., T. D. Kennedy, E. Dratewka-Kos, and E. G. Lane. 1990. Covalently bonded and adventitious glycans in germin. J. Biol. Chem. 265: 12503–12512.
136. Jain, S., and D. E. Ohman. 1998. Deletion of *algK* in mucoid *Pseudomonas aeruginosa* blocks alginate polymer formation and results in uronic acid secretion. J. Bacteriol. 180:634–641.
137. Jakoby, M., M. Tesch, H. Sahm, R. Kraemer, and A. Burkovski. 1997. Isolation of the *Corynebacterium glutamicum glnA* gene encoding glutamine synthase I. FEBS Microbiol. Lett. 154:81–88.
138. Jamaux, I., B. Gelie, and C. Lamarque. 1995. Early stages of infection of rapeseed petals and leaves by *Sclerotinia sclerotiorum* revealed by scanning electron microscope. Plant Pathol. 44:22–30.
139. Jensen, S. O., and P. R. Reeves. 1998. Domain organisation in phosphomannose isomerases (types I and II). Biochim. Biophys. Acta 1382:5–7.
140. Jiang, X. M., B. Neal, F. Santiago, S. J. Lee, L. K. Romana, and P. R. Reeves. 1991. Structure and sequence of the *rfb* (O antigen) gene cluster of *Salmonella* serovar *typhimurium* (strain LT2). Mol. Microbiol. 5:695–713.
141. Jones, A. M. 1994. Auxin-binding proteins. Annu. Rev. Plant Physiol. Plant Mol. Biol. 45:393–420.
142. Jones, A. M., K. H. Im, M. A. Savka, M. J. Wu, N. G. DeWitt, R. Shillito, and A. N. Binns. 1998. Auxin-dependent cell expansion mediated by overexpressed auxin-binding protein. Science 282:1114–1117.
143. Jones, D. C., and R. A. Cooper. 1990. Catabolism of 3-hydroxybenzoate by the gentisate pathway in *Klebsiella pneumoniae* M5a1. Arch. Microbiol. 154:489–495.
144. Jump, J. A. 1954. Studies on sclerotization in *Physarum polycephalum*. Am. J. Bot. 41:561–567.
145. Kaneko, T., and S. Tabata. 1997. Complete genome structure of the unicellular cyanobacterium *Synechocystis* sp. PCC6803. Plant Cell Physiol. 38: 1171–1176.
146. Kaneko, T., S. Sato, H. Kotani, A. Tanaka, E. Asamizu, Y. Nakamura, N. Miyajima, M. Hirosawa, M. Sigiura, S. Sasamoto, T. Kimura, T. Hosouchi, A. Matsuno, A. Muraki, N. Nakazaki, K. Naruo, S. Okumura, S. Shimpo, C. Takeuchi, T. Wada, A. Watanabe, M. Yamada, M. Yasuda, and S. Tabata. 1996. Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. DNA Res. 3:109–136.
147. Kawarabayasi, Y., M. Sawada, H. Horikawa, Y. Haikawa, Y. Hino, S. Yamamoto, M. Sekine, S. Baba, H. Kosugi, A. Hosoyama, Y. Nagai, M. Sakai, K. Ogura, R. Otuka, H. Nakazawa, M. Takamiya, Y. Ohfuku, T. Funahashi, T. Tanaka, Y. Kudoh, J. Yamazaki, N. Kushida, A. Oguchi, K. Aoki, Y. Nakamura, T. F. Ruhh, K. Horikoshi, Y. Masuchi, H. Shizuya, and H. Kikuchi. 1998. Complete sequence and gene organization of the genome of a hyper-thermophilic archaebacterium, *Pyrococcus horikoshii* OT3. DNA Res. 5:55–76.
148. Kelemen, G. H., P. Brian, K. Flärdh, L. Chamberlin, K. F. Chater, and M. J. Buttner. 1998. Developmental regulation of transcription of *whiE*, a locus specifying the polyketide spore pigment in *Streptomyces coelicolor* A3(2). J. Bacteriol. 180:2515–2521.
149. Kemper, M. J., D. Nolkemper, X. Rogiers, K. Timmermann, E. Sturm, M. Malago, C. E. Broelsch, M. Burdelski, and D. E. Muller-Wiefel. 1998. Preemptive liver transplantation in primary hyperoxaluria type 1: timing and preliminary results. J. Nephrol. 11:46–48.
150. Keyhani, N. O., and S. Roseman. 1997. Wild-type *Escherichia coli* grows on the chitin disaccharide, *N,N'*-diacetylchitobiose, by expressing the *cel* operon. Proc. Natl. Acad. Sci. USA 94:14367–14371.
151. Kiemer, P., B. Tshisuaka, S. Fetzner, and F. Lingens. 1996. Degradation of benzoate via benzoyl-coenzyme A and gentisate by *Bacillus stearothermophilus* PK1, and purification of gentisate 1,2-dioxygenase. Biol. Fertil. Soils 23:307–313.
152. Kim, H-Y., D. Schlictman, S. Shankar, Z. Xie, A. M. Chakrabarty, and A. Kornberg. 1998. Alginate, inorganic polyphosphate, GTP and ppGpp synthesis co-regulated in *Pseudomonas aeruginosa*: implications for stationary phase survival and synthesis of RNA/DNA precursors. Mol. Microbiol. 27:717–725.
153. Klenk, H. P., R. R. Clayton, J. F. Tomb, O. White, K. E. Nelson, K. A. Ketchum, R. J. Dodson, M. Gwinn, E. K. Hickey, J. D. Peterson, D. L. Richardson, A. R. Kerlavage, D. E. Graham, N. C. Kyrpides, R. D. Fleischmann, J. Quackenbush, N. H. Lee, G. G. Sutton, S. Gill, E. F. Kirkness, B. A. Dougherty, K. McKenney, M. D. Adams, B. Loftus, S. Peterson, C. I. Reich, L. K. McNeil, J. H. Badger, A. Glodek, L. Zhou, R. Overbeek, J. D. Gocayne, J. F. Weidman, L. McDonald, T. Utterback, M. D. Cotton, T. Spriggs, P. Artiach, B. P. Kaine, S. M. Sykes, P. W. Sadow, K. P. D'Andrea, C. Bowman, C. Fujii, S. A. Garland, T. M. Mason, G. J. Olsen, C. M. Fraser, H. O. Smith, C. R. Woese, and J. C. Venter. 1997. The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. Nature 390:364–370.
154. Knutson, D. M., A. S. Hutchins, and K. Cromack. 1980. The association of calcium oxalate-utilizing *Streptomyces* with conifer ectomycorrhizae. Antonie Leeuwenhoek 46:611–619.
155. Ko, T-P., J. D. Ng, and A. McPherson. 1993. The three-dimensional structure of canavalin from jack bean (*Canavalia ensiformis*). Plant Physiol. 101:729–744.
156. Reference deleted.
157. Koivula, T. T., H. Hemilä, R. Pakkanen, M. Sibakov, and I. Palva. 1993 Cloning and sequencing of a gene encoding acidophilic amylase from *Bacillus acidocaldarius*. J. Gen. Microbiol. 139:2399–2407.
158. Koonin, E. V., A. R. Mushegian, M. Y. Galperin, and D. R. Walker. 1997. Comparison of archaeal and bacterial genomes: computer analysis of protein sequences predicts novel functions and suggests a chimeric origin for the archaea. Mol. Microbiol. 25:619–637.
159. Koonin, E. V., R. L. Tatusov, and M. Y. Galperin. 1998. Beyond complete genomes: from sequence to structure and function. Curr. Opin. Struct. Biol. 8:355–363.
160. Köplin, R., G. Wang, B. Hötte, U. B. Priefer, and A. Pühler. 1993. A 3.9-kb DNA region of *Xanthomonas campestris* pv. *campestris* that is necessary for lipopolysaccharide production encodes a set of enzymes involved in the synthesis of dTDP-rhamnose. J. Bacteriol. 175:7786–7792.
161. Kotsira, V. P., and Y. D. Clonis. 1997. Oxalate oxidase from barley roots: purification to homogeneity and study of some molecular, catalytic, and binding properties. Arch. Biochem. Biophys. 340:239–249.
162. Kotsira, V. P., and Y. D. Clonis. 1998. Chemical modification of barley root oxalate oxidase shows the presence of a lysine, a carboxylate, and disulphides, essential for enzyme activity. Arch. Biochem. Biophys. 356:117–126.
163. Kuan, L-C., and M. Tien. 1993. Stimulation of Mn peroxidase activity: A possible role for oxalate in lignin biodegradation. Proc. Natl. Acad. Sci. USA 90:1242–1246.
164. Kucharczyk, R., M. Zagulski, J. Rytka, and C. J. Herbert. 1998. The yeast gene *YJR025c* encodes a 3-hydroxyanthranilic acid dioxygenase and is involved in nicotinic acid biosynthesis. FEBS Lett. 424:127–130.
165. Kuhner, C. H., P. A. Hartman, and M. J. Allison. 1996. Generation of a proton motive force by the anaerobic oxalate-degrading bacterium *Oxalobacter formigenes*. Appl. Environ. Microbiol. 62:2494–2500.
166. Kunst, F., N. Ogasawara, I. Moszer, A. M. Albertini, G. Alloni, V. Azevedo, M. G. Bertero, P. Bessieres, A. Bolotin, S. Borchert, R. Borriss, L. Boursier, A. Brans, M. Braun, S. C. Brignell, S. Bron, S. Brouillet, C. V. Bruschi, B.

, V. Capunno, N. M. Carter, S. K. Choi, J. J. Codoni, I. F. Connell, N. J. Cummings, R. A. Daniel, F. Denizot, K. M. Devine, A. Dusterhoft, S. D. Ehrlich, P. T. Emmerson, K. D. Entian, J. Errington, C. Fabret, E. Ferrari, D. Foulger, C. Fritz, M. Fujita, Y. Fujita, S. Fuma, A. Galizzi, N. Galleron, S. Y. Ghim, P. Glaser, A. Goffeau, E. J. Golightly, G. Grandi, G. Guiseppi, B. J. Guy, K. Haga, J. Haiech, C. R. Harwood, A. Hénaut, H. Hilbert, S. Holsappel, S. Hosono, M. F. Hullo, M. Itaya, L. Jones, B. Joris, D. Karamata, Y. Kasahara, M. Klaerr-Blanchard, C. Klein, Y. Kobayashi, P. Koetter, G. Koningstein, S. Krogh, M. Kumano, K. Kurita, A. Lapidus, S. Lardinois, J. Lauber, V. Lazarevic, S. M. Lee, A. Levine, H. Liu, S. Masuda, C. Mauel, C. Medigue, N. Medina, R. P. Mellado, M. Mizuno, D. Moestl, S. Nakai, M. Noback, D. Noone, M. O'Reilly, K. Ogasawara, N. Ogiwara, B. Oudega, S. H. Park, V. Parro, T. M. Pohl, D. Portetelle, S. Porwollik, A. M. Prescott, E. Presecan, P. Pujic, B. Purnelle, G. Rapoport, M. Rey, S. Reynolds, M. Rieger, C. Rivolta, E. Rocha, B. Roche, M. Rose, Y. Sadaie, T. Sato, E. Scanlan, S. Schleich, R. Schroeter, F. Scoffone, J. Sekiguchi, A. Sekowska, S. J. Seror, P. Serror, B. S. Shin, B. Soldo, A. Sorokin, E. Tacconi, T. Takagi, H. Takahashi, K. Takemaru, M. Takeuchi, A. Tauch, M. Tokuyoshi, T. Tanaka, P. Terpstra, A. Tognoni, V. Tosato, S. Uchiyama, M. Vandenbol, F. Vannier, A. Vassarotti, A. Viari, R. Wambutt, E. Wedler, H. Wedler, T. Weitzenegger, P. Winters, A. Wipat, H. Yamamoto, K. Yamane, K. Yasumoto, K. Yata, K. Yoshida, H. F. Yoshikawa, E. Zumstein, H. Yoshikawa, and A. Danchin. 1997. The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. Nature 390:249–256.

mkis, L., D. H. Anderson, and A. J. Duncan. 1998. Conditioned flavor aversions in sheep: the relationship between the dose rate of a secondary plant compound and the acquisition and persistence of aversions. Nutr. 79:55–62.

mer, A. G., and A. R. Fersht. 1997. Glutamine, alanine or glycine ts inserted into the loop of a protein have minimal effects on stability olding rates. J. Mol. Biol. 273:330–337.

ry, M. M., and C. W. Parkins. 1993. Calcium oxalate crystal deposition crotizing otomycosis caused by *Aspergillus niger*. Mod. Pathol. 6:493–

, B. G. 1994. Oxalate, germin and the extracellular matrix. FASEB J. –301.

, B. G., F. Bernier, E. Dratewka-Kos, R. Shafai, T. D. Kennedy, C. J. R. Munro, T. Vaughan, D. Walters, and F. Altomare. 1991. Homologies between members of the germin gene family in hexaploid wheat similarities between these wheat germins and certain *Physarum* spherudes. J. Biol. Chem. 266:10461–10469.

, B. G., A. C. Cuming, J. Frégeau, N. C. Carpita, W. J. Hurkman, F. ier, E. Dratewka-Kos, and T. D. Kennedy. 1992. Germin isoforms are rete temporal markers of wheat development. Eur. J. Biochem. 209: –969.

e, B. G., J. M. Dunwell, J. Ray, M. R. Schmitt, and A. C. Cuming. 1993. rmin, a marker of early plant development, is an oxalate oxidase. J. Biol. m. 268:12239–12242.

ie, B. G., Z. F. Grzelczak, T. D. Kennedy, R. Kajioka, J. Orr, S. gostino, and A. Jaikaran. 1986. Germin. Compartmentation of two ms of the protein by washing growing wheat embryos. Biochem. Cell l. 64:1025–1037.

agman, L. J., and L. C. Allen. 1998. An enzymatic method for oxalate tomated using the Hitachi 911 analyzer. Clin. Biochem. 31:429–432.

thika, K. M., S. Sharma, K. V. Inamdar, and K. G. Raghavan. 1995. alate depletion from leafy vegetables using alginate entrapped banana alate oxidase. Biotechnol. Lett. 17:407–410.

wrence, M. C., T. Izard, M. Beuchat, R. J. Blagrove, and P. M. Colman. 94. Structure of phaseolin at 2.2 Å resolution. Implications for a common silin/legumin structure and the genetic engineering of seed storage proins. J. Mol. Biol. 238:748–776.

thel, C., D. A. Los, H. Wada, J. Gyorgyei, I. Horvath, E. Kovacs, N. luratu, and L. Vigh. 1993. A second *groEL*-like gene, organised in a *vESL* operon is present in the genome of Synechocystis sp. PCC 6803. Biol. Chem. 268:1799–1804.

eitao, J. H., and I. Sa-Correia. 1997. Oxygen-dependent upregulation of ranscription of alginate genes *algA*, *algC* and *algD* in *Pseudomonas aeruginosa*. Res. Microbiol. 148:37–43.

eltner, A., E. Jensen-Jarolim, R. Grimm, B. Wüthrich, H. Ebner, O. cheiner, D. Kraft, and C. Ebner. 1998. Allergens in pepper and paprika. mmunologic investigation of the celery-birch-mugwort-spice syndrome. Allergy 53:36–41.

Li, Y. 1995. Structure and function analysis of the *usaA*, *ptpA*, and *cysA* loci of *Streptomyces coelicolor*. Ph.D. thesis. Ohio State University, Columbus.

Li, Y., and W. R. Strohl. 1996. Cloning, purification, and properties of a phosphotyrosine protein from *Streptomyces coelicolor* A3(2). J. Bacteriol. 178:136–142.

Lin, T. J., D. Z. Hung, W. H. Hu, D. Y. Yang, T. C. Wu, and J. F. Deng. 1998. Calcium oxalate is the main toxic component in clinical presentations of *Alocasia macrorrhiza* (L) Schott and Endl poisonings. Vet. Hum. Toxicol. 40:93–95.

Liu, H. W., and J. S. Thorson. 1994. Pathways and mechanisms in the biogenesis of novel deoxysugars by bacteria. Annu. Rev. Microbiol. 48:223–256.

185. Loewus, F. A., K. Saito, R. K. Sato, and E. Maring. 1995. Conversion of 2-deoxy-threo-D-ery-throascorbic acid and oxalic acid in *Sclerotinia sclerotiorum*. Biochem. Biophys. Res. Commun. 212:196–203.

186. Lung, H.-Y., A. L. Baetz, and A. B. Peck. 1994. Molecular cloning, DNA sequence, and gene expression of the oxalyl-coenzyme A decarboxylase gene, *oxc*, from the bacterium *Oxalobacter formigenes*. J. Bacteriol. 176: 2468–2472.

187. Lung, H.-Y., J. G. Cornelius, and A. B. Peck. 1991. Cloning and expression of the oxalyl-CoA decarboxylase gene from the bacterium, *Oxalobacter formigenes*: prospects for gene therapy to control Ca-oxalate kidney stone formation. Am. J. Kidney Dis. 17:381–385.

188. Ma, J. F., S. Hiradate, and H. Matsumoto. 1998. High aluminum resistance in buckwheat. II. Oxalic acid detoxifies aluminum internally. Plant Physiol. 117:753–759.

189. Macpherson, D. F., P. A. Manning, and R. Morona. 1994. Characterisation of the Dtdp-rhamnose biosynthetic genes encoded in the *rfb* locus of *Shigella flexneri*. Mol. Microbiol. 11:281–292.

190. Malherbe, P., C. Kohler, M. Da Prada, G. Lang, V. Kiefer, R. Schwartz, H. W. Lahm, and A. M. Cesura. 1994. Molecular cloning and functional expression of human 3-hydroxyanthranilic-acid dioxygenase. J. Biol. Chem. 269:13792–13797.

191. Mans, K. J., C. D. Ratich, and L. McFetridge. July 1998. Materials and methods for detecting oxalate. U.S. patent 5,776,701.

192. Marciano, P., P. Di Lenna, and P. Marga. 1983. Oxalic acid, cell wall degrading enzymes and their significance in the virulence of two *Sclerotinia sclerotiorum* isolates on sunflower. Plant Pathol. 22:339–345.

193. Maroida, C. L., and M. A. Valvano. 1995. Genetic analysis of the dTDP-rhamnose biosynthesis region of the *Escherichia coli* VW187 (O7:K1) *rfb* gene cluster: identification of functional homologs of *rfbB* and *rfbA* in the *rff* cluster and correct location of the *rffE* gene. J. Bacteriol. 177:5539–5546.

194. Marsden, A. F., B. Wilkinson, J. Cortes, N. J. Dunster, J. Staunton, and P. F. Leadlay. 1998. Engineering broader specificity into an antibiotic-producing polyketide synthase. Science 279:199–202.

195. Masirevic, S., and T. J. Gulya. 1992. *Sclerotinia* and *Phomopsis*, two devastating sunflower pathogens. Field Crop. Res. 30:271–300.

196. May, T. B., D. Shinnarger, A. Boyd, and A. M. Chakrabarty. 1994. Identification of amino acid residues involved in the activity of phosphomannose isomerase-guanosine 5′-diphospho-D-mannose pyrophosphorylase. J. Biol. Chem. 269:4872–4877.

197. McCubbin, W. C., C. M. Kay, T. D. Kennedy, and B. G. Lane. 1987. Germin: physicochemical properties of the glycoprotein which signals the onset of growth in the germinating wheat embryo. Biochem. Cell Biol. 65:1039–1048.

198. Mehta, A., and A. Datta. 1991. Oxalate decarboxylase from *Collybia velutipes*. Purification, characterisation and cloning. J. Biol. Chem. 266:23548–23553.

199. Mehta, A., K. Naturajan, A. Rainu, and A. Datta. 1994. A step towards developing transgenic plants with high nutritional quality. Proc. Indian Natl. Sci. Acad. B60:375–380.

200. Membré, N., A. Berna, G. Neutelings, A. David, H. David, D. Staiger, J. S. Vásquez, M. Raynal, M. Delseny, and F. Bernier. 1997. cDNA sequence, genome organization and differential expression of three *Arabidopsis* genes for germin/oxalate oxidase-like proteins. Plant Mol. Biol. 35:459–469.

201. Mester, T., and J. A. Field. 1998. Characterization of a novel manganese peroxidase-lignin peroxidase hybrid isozyme produced by *Bjerkandera* species strain BOS55 in the absence of manganese. J. Biol. Chem. 19:15412–15417.

202. Micales, J. A. 1995. *In vitro* oxalic acid production by the brown-rot fungus *Postia placenta*. Mater. Org. 29:159–176.

203. Micales, J. A. 1997. Localization and induction of oxalate decarboxylase in the brown-rot wood decay fungus *Postia placenta*. Int. Biodeterior. Biodegrad. 39:125–132.

204. Michalowski, C. B., and H. J. Bohnert. 1992. Nucleotide sequence of a root-specific transcript encoding a germin-like protein from the halophyte *Mesembryanthemum crystallinum*. Plant Physiol. 100:537–538.

205. Mitchison, M., D. M. Bulach, T. Vinh, K. Rajakumar, S. Faine, and B. Adler. 1997. Identification and characterisation of the dTDP-rhamnose biosynthesis and transfer genes of the lipopolysaccharide-related *rfb* locus in *Leptospira interrogans* serovar Copenhageni. J. Bacteriol. 179:1262–1267.

206. Monte, J. V., C. Casanova, and C. Soler. 1998. Presence and organization of an osmotic stress response domain in wild *Triticeae* species. Theor. Appl. Genet. 96:76–79.

207. Moon, Y. H., S. Chae, J. Y. Jung, and G. An. 1998. Expressed sequence tags of radish flower buds and characterization of a *CONSTANS LIKE 1* gene. Mol. Cells 8:452–458.

208. Moreira, M. T., C. Feijoo, T. Mester, P. Moyorga, R. Sierra-Alvarez, and J. A. Field. 1998. Role of organic acids in the manganese-independent biobleaching system of *Bjerkandera* sp. strain BOS55. Appl. Environ. Microbiol. 64:2409–2417.

209. Morton, R. A., M. J. Dilworth, and B. Wiemecke. May 1994. Biological